Identifying correlates of CTL-mediated control of HIV using computational and biomedical approaches

by

Eric Martin

B.Sc., Simon Fraser University, 2009

Thesis Submitted In Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Spring 2013

Approval

Name: **Eric Martin**

Degree: **Master of Science (Molecular Biology & Biochemistry)**

Title of Thesis: ***Identifying correlates of CTL-mediated control of HIV using computational and biomedical approaches***

Examining Committee: **Chair:** Peter Unrau
Associate Professor

Mark Brockman
Senior Supervisor
Associate Professor

Zabrina Brumme
Co-Supervisor
Assistant Professor

Jonathan Choy
Committee Member
Assistant Professor

Robert Holt
Committee Member
Associate Professor

Fiona Brinkman
Internal Examiner
Professor, Molecular Biology
Simon Fraser University

Date Defended/Approved: January 8, 2013

Partial Copyright Licence

Abstract

Cytotoxic T lymphocytes (CTL) select HIV variants that encode mutations in peptide epitopes presented by HLA. Studying HLA-associated viral polymorphisms may therefore elucidate characteristics of effective CTL responses. Using population-level data, we identified 2100 HLA-associated polymorphisms located throughout the HIV proteome. In 9-mer epitopes, viral mutations at HLA anchor residues occurred two-fold more frequently than expected and reduced predicted peptide-HLA binding affinity by ten-fold. Epitopes presented by more protective HLA displayed greater enrichment of anchor residue mutations. CTL epitope recognition is determined by its T-cell receptor (TCR). To assess the impact of epitope mutations on TCR signaling, we implemented an *in vitro* reporter cell assay. This method was validated using a TCR specific for the HLA-A*02:01-FK10 peptide (FLGKIWPSYK), and proved sensitive, scalable, and robust. Alanine mutants of the TCR and peptide variants identified an interaction between FK10 position 4 and TCR alpha CDR3 that may be critical for signaling.

Keywords: HIV-1; Human leukocyte antigen; Cytotoxic T lymphocyte; T-cell receptor; Immune control

Dedication

To my wife, Aiste Guden. I love you very much. To my parents, John and Donna, my supervisors, Zabrina and Mark, my lab partners, Gursev, Tristan, Tallie, Ryan, Anh, Laura, Phillip, Bemulu, and Aniqa. You are all great.

Acknowledgements

We thank Dr. Brad Jones and Dr. Mario Ostrowski for providing the 5B2 CTL cell line. We thank Dr. Simon Mallal and Dr. Mina John for providing sample from the ACTG cohorts, and Dr. Richard Harrigan for providing data on the HOMER cohort.

Dr. Christian Hofmann for providing advice on the TCR assay.

Table of Contents

List of Tables

List of Figures

List of Acronyms

APC	Antigen-presenting cell
CDR	Complementarity determining region
CTL	Cytotoxic T-lymphocyte
ELISPOT	Enzyme-linked immunosorbent spot
IFNg	Interferon-gamma
ITAM	Immunoreceptor tyrosine-based activation motif
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
NF-AT	Nuclear factor of activated T-cells
PBMC	Peripheral blood mononuclear cell
pHLA	Peptide-HLA
SPR	Surface plasmon resonance
TCR	T-cell receptor

Executive Summary

Host expression of Human Leukocyte Antigen (HLA) class I alleles is associated with clinical outcome following human immunodeficiency virus (HIV) infection, indicating that CD8+ cytotoxic T lymphocytes (CTL) play a critical role in reducing viral pathogenesis. However, mechanisms of CTL-mediated immune control remain unclear.

HIV evades CTL responses through the selection of HLA-associated polymorphisms located within or near targeted viral epitopes. While patterns of CTL escape mutations can differ widely among individuals, they appear to be largely consistent within the context of particular host HLA alleles at the population level.

Because HLA-associated polymorphisms define how HIV evades host CTL, they provide a measure of the *in vivo* characteristics associated with CTL responses that place selective pressure on the virus. Therefore, we hypothesized that a better understanding of the patterns of HLA-associated mutations in HIV-infected individuals would help to clarify immune parameters that constitute an effective CTL response. To address this, we analyzed linked viral sequences and HLA genotype data collected from a cohort of N=1800 chronic HIV-1 subtype B infected individuals. We identified over 2100 HLA-associated polymorphisms, located at 804 viral codons. We next assessed whether CTL escape mutations developed at preferred sites within known HLA-restricted CTL epitopes in the virus proteome. Among 9-mer epitopes, we observed that mutations located at HLA anchor residues (typically peptide positions 2 and 9) occurred two-fold more frequently than expected by chance, and they were computationally predicted to reduce peptide-HLA binding by approximately ten-fold. Known CTL epitopes presented by protective HLA alleles (defined using hazard ratios for progression to AIDS) were enriched for mutations at anchor residues, suggesting that other changes in the epitope were less able to evade CTL. This result is consistent with increased CTL breadth and cross-reactivity for viral variants in the case of protective HLA alleles.

The ability of a T-cell to recognize virus-infected cells is determined by its antigen-specific T-cell receptor (TCR). The TCR repertoire in an individual is highly diverse and multiple TCR typically recognize a foreign peptide antigen, resulting in an oligo-clonal response. Extensive sequence diversity is observed among anti-HIV TCR;

however, there is very little understanding of their functional diversity, particularly as it relates to recognition of viral mutations that contribute to CTL escape. To examine this, we implemented an *in vitro* reporter T-cell assay to measure TCR signalling in response to peptide-HLA ligands, allowing us to directly assess the impact of escape variants on TCR recognition. This assay was validated using a human TCR specific for the HLA-A*02:01 restricted, HIV Gag FK10 peptide, and it proved to be sensitive, scalable, and robust. Using alanine substitution, we determined that peptide positions 1, 2, 4, 6, and 7 were necessary for TCR signalling. Bioinformatics analysis suggested that positions 1, 2, and 7 participate in HLA binding; therefore, we predict that position 4 and position 6 of the epitope is a TCR contact. Consistent with this, several naturally occurring viral polymorphisms at position 9 (H441X) did not affect TCR signalling, but a K436R mutation at position 4 abrogated TCR function. This *in vitro* assay is amenable to molecular manipulation, and alanine-scanning mutagenesis of the complementarity determining region 3 (CDR3) on the TCR alpha and beta chains identified numerous positions that were necessary for peptide-specific TCR signalling. Using a structural model of this TCR, we identified a putative electrostatic interaction between a Glutamic acid at position 4 on the TCR alpha CDR3 and a Lysine (K436) at position 4 on the FK10 peptide. Future approaches that combine bioinformatics analysis of HIV evolution, peptide/HLA binding analyses, and experimental methods to measure TCR function will allow the mechanisms of CTL-mediated control and viral escape to be examined more systematically. This work will result in a better understanding of effective antiviral immune responses and influence HIV vaccine design strategies.

1. HIV and the T-cell response to infection: a review

1.1. HIV: virus and pathogenesis

Human immunodeficiency virus type 1 (HIV) is a retrovirus that infects CD4+ T-cells. First identified in 1983[1], HIV is recognized along with tuberculosis and malaria as major global health challenges[2]. The single-stranded RNA genome of HIV encodes 9 genes and 15 proteins[3], including the common retroviral genes *gag*, *pol*, and *env*, and the regulatory and accessory genes *vif*, *vpr*, *vpu*, *rev*, *tat*, and *nef* (Figure 1.1).

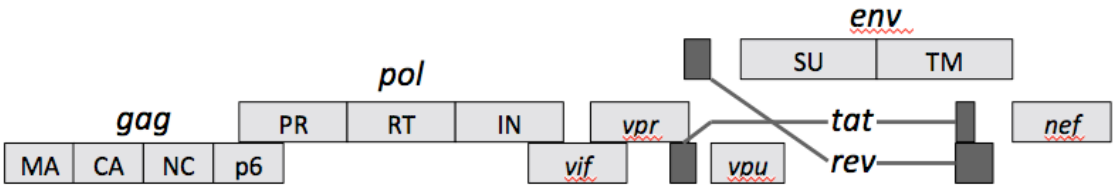

Figure 1.1 The HIV-1 genome.

The HIV-1 genome contains 9 open reading frames, common retroviral genes Gag, Pol, and Nef, and regulatory genes vif, vpr, vpu, rev, tat, and nef.

The *gag*, *pol*, and *env* genes are common to all retroviruses and critical for the structure and life cycle of the virus[3]. The Gag polyprotein is cleaved into matrix (MA), capsid (CA), nucleocapsid (NC), and p6, and is necessary for the proper structural formation of the virion. MA targets viral proteins to the host membrane, CA forms an oligomeric inner protein core, and NC directs full-length viral RNA for assembly into a virion. The envelope (Env) polyprotein is cleaved to form a surface (SU) glycoprotein gp120 and transmembrane (TM) subunit gp41, which form a trimer of gp120/gp41 heterodimers and determines viral tropism. Gp120 binds to CD4, the viral receptor on target T-cells, and a co-receptor protein, either CCR5[4,5] or CXCR4[6]. GP41 subsequently mediates fusion between the viral membrane and target T-cell membrane. The Pol polyprotein is cleaved into the enzymatically active protease (PR), reverse transcriptase

(RT), and integrase (IN) subunits. PR is essential for virion maturation through the cleavage of polyproteins into active subunits. RT contains RNA and DNA dependent polymerase activity, and along with an RNase H domain, converts the viral RNA genome into a DNA form that can be integrated into the host genome. IN then facilitates the integration of viral DNA into the host genome as a provirus. The regulatory and accessory genes *vif, vpr, vpu, rev, tat*, and *nef* are not found in other classes of retroviruses, but are crucial to facilitate the *in vivo* life cycle of HIV, to circumvent innate and adaptive host immune factors, and to modulate the establishment and maintenance of viral latency[7]. Of particular interest is the ability of the Nef protein to downregulate host Human Leukocyte Antigen (HLA) class I, which attenuates host CTL responses[8]. For a detailed review of different HIV protein functions, please read[3].

Gradual loss of CD4+ T-cells as a result of HIV infection impairs cellular immunity and increases morbidity and mortality associated with normally benign opportunistic infections and cancers, which is now referred to as acquired immunodeficiency syndrome (AIDS)[9]. The prevalence of HIV-1 world-wide in 2009 was 0.8%; however, populations from sub-Saharan Africa suffer 67% of all cases, amounting to ~24 million of 35 million persons living with HIV globally[10]. HIV is also a substantial health burden in Canada, where, as of 2009, 65,000 people live with HIV, and one third of individuals remain unaware of their infection status[11]. Improved HIV surveillance and treatment strategies are needed, especially among aboriginal peoples and injection drug users, who continue to have higher HIV incidence and prevalence than other North American populations[12].

1.2. Highly active antiretroviral therapy (HAART)

Highly active antiretroviral therapy (HAART) is the major HIV research and clinical success from the past 30 years[13,14]. HAART can suppress viremia to undetectable levels (less than 50 RNA copies/mL in plasma), lead to recovery in CD4 T-cell function[15] and significantly extend the life of HIV+ individuals. In addition to benefiting the individual, because higher plasma viremia is associated with a greater risk factor for HIV transmission[16], public health strategies to reduce the average community viral load through enhanced availability of HAART may also be successful to decrease HIV incidence[17]. However, due to the ability of HIV to establish stable reservoirs of latent

infection[18], suspending drug treatment results in a rapid rebound in viral load[19]. Thus, HIV infection currently requires lifelong therapy.

Low fidelity of the HIV-1 RT enzyme, which has an average error rate of 1 in 1700 nucleotides in each newly synthesized viral DNA [20], results in rapid viral mutation. This allows the HIV proteome to accrue mutations that can mediate drug resistance[21], disrupt antibody recognition[22], and disrupt T-cell recognition[23]. The use of multi-drug cocktails have proven necessary to create a sufficient genetic barrier to prevent drug resistance mutations from occurring; however, high drug burden is associated with side effects such as renal toxicity[24], representing a barrier to treatment adherence. The high cost of HAART and requirement for frequent clinical monitoring are additional challenges to widespread use of therapy, particularly in developing regions of the world most affected by HIV/AIDS. Therefore, development of a safe and effective HIV vaccine remains a high priority of the research community.

1.3. Host immune responses to HIV infection

Although advances in virological characterizations of HIV have led to successful drug therapies, studying the natural immune response to HIV infection is essential to understand the pathogenesis of this disease and may allow the development of an HIV vaccine. Thus, much attention has been put towards host/HIV immune interactions.

1.3.1. The immune response: overview

The immune system can be classified into an innate and an adaptive response. The innate response is characterized by recognition of pathogen associated molecular patterns (PAMPs); for example, detection of bacterial lipopolysaccharides (LPS) can result in inflammation and recruitment of immune cells to the site of infection, and detection of viral RNA species in the cytoplasm of infected cells can lead to production of type 1 interferon and proinflammatory cytokines[25]. Although innate responses do not confer immunological memory, they are rapid, they can independently control some infections, and they initiate critical steps leading to an adaptive response[26]. The adaptive response provides specific recognition of pathogen-derived protein epitopes and establishes immunological memory. An adaptive response begins when a tissue-

resident dendritic cell is activated by a PAMP directly, or through the consequent inflammation. Activated DCs travel to regional lymph nodes and activate T-cells and B-cells to differentiate in a manner tailored to the type of pathogen encountered. B-cells secrete antibodies, which can then proceed to bind soluble antigen, or to a surface such as an infected cell or parasite. Alternatively, T-cells represent cellular components of the immune system, which are necessary to recognize intracellular antigens and pathogen-infected cells.

T-cells are crucial for the development of an adaptive immune response. They respond to intracellular peptide antigens presented on the cell surface in complex with HLA. T-cells are defined by the presence of a T-cell receptor (TCR) on the cell surface, which determines antigen specificity. Two classes of T-cells are the CD4-expressing T-helper cell, and the CD8-expressing cytotoxic T-lymphocyte (CTL). CD4+ helper T-cells are immune-modulatory cells, which recruit appropriate downstream immune responses, depending on the cytokine environment[27]. Intracellular pathogens induce CD4+ T-cells to differentiate into the T_H1 lineage, which activate CTL and macrophages. CTL are crucial in clearance of intracellular pathogen by identifying and killing pathogen infected cells through direct and indirect cytolytic mechanisms. As T-cell activation causes cell division and cell death, the totality of different T-cells in the body – referred to as the T-cell repertoire – changes over time due to antigen-induced clonal expansion and deletion.

1.3.2. Spontaneous control of HIV offers clues to a vaccine

The rate of HIV disease progression varies widely following infection. Some individuals display a immediate loss of CD4 T-cells and rapid progression to AIDS while other display slower rates of CD4 decline and a relatively non-progressive disease course[28]. Approximately 1 in 300 HIV+ individuals - called elite controllers – are able to spontaneously suppress HIV infection to undetectable levels in the blood (less than 50 RNA copies/mL) without drug therapy[29]. It is thought that differences in the antiviral immune response of HIV long-term non-progressors and elite controllers contributes to their phenotype, and therefore designing a vaccine that can mimic these protective immune qualities remains a goal in HIV research. Notably, certain HLA class I alleles, such as B*57 and B*27, are associated with different rates of progression to AIDS[30], and these same HLA alleles are enriched among spontaneous HIV controllers[31]. Indeed, population level studies have quantified this protective effect by calculating the relative

risk (or, *hazard ratio*) of developing AIDS among untreated patients expressing a particular HLA allele[32]. HLA alleles can shape the qualities of a T-cell response, and these differences in hazard ratio presumably serve as markers for an immunological correlate of CTL-mediated protection.

1.3.3. The CTL response is a correlate of HIV virologic control

CTL responses are positively correlated with the clinical outcome of HIV infection. Within the first few months after HIV infection, plasma viremia typically declines and stabilizes at a set point viral load. Individuals with a high set-point viral load are at greatest risk for developing AIDS[33]. Several lines of evidence suggest that CTL are important in influencing set-point viremia. First, CTL responses are temporally associated with the establishment of viral load set point[34]. Second, certain HLA-I alleles, which present viral epitopes to CTL, are associated with rapid progression to AIDS, while others are enriched in elite controllers[35]. Third, in the rhesus macaque model of SIV infection, transient depletion of CD8+ cells resulted in a rapid increase in plasma viremia, which was subsequently suppressed as CD8 cells recovered in vivo[36]. Finally, a large number of HIV mutations are associated with specific host HLA class I alleles, suggesting that CTL pressure is a major determinant of viral evolution *in vivo* [37]. These data indicate that a particularly effective CTL response might contain viral infection. However, the antiviral mechanisms of CTL-mediated control need to be understood before the development of an HIV vaccine is feasible.

1.4. Mechanics of the CTL response

Antiviral CTL responses require processing of intracellular antigens to generate short immunogenic peptides, presentation of these peptides by HLA class I molecules on the infected cell surface, and recognition of the peptide-HLA complex (pHLA) by circulating CTL via their antigen-specific TCR.

1.4.1. Antigen processing

The preparation of cytosolic peptides – known as antigen processing – and the display of these peptides on the infected cell surface – known as antigen presentation – are key components of a T-cell response.

During antigen processing, cytosolic proteins are cleaved by the proteasome into peptide fragments, which are imported into the endoplasmic reticulum (ER) by the Transporter Associated with Antigen Processing (TAP) and trimmed to short 8-11mer peptides by ER-resident peptidases. Proteasomal cleavage defines the C-terminus of the epitope, leaving N-terminally extended products to be trimmed by peptidases[38]. The proteasome, importer, and peptidases display sequence specificity, such that a large, yet finite, number of correctly processed antigenic peptides are produced[39,40].

1.4.2. Presentation on HLA class I

Peptides in the ER are next loaded onto host HLA class I molecules for presentation at the cell surface. The peptide side-chains interact with spatial grooves in between the α_1 and α_2 domains of the HLA heavy chain protein. The peptide-specificity of HLA class I proteins are determined by the HLA heavy chain, encoded by three polymorphic genes: HLA-A, HLA-B, and HLA-C[41]. Due to the high degree of HLA diversity, individuals have different HLA alleles, which present different peptide repertoires. Typically, for a 9-mer peptide sequence, side chains of position 2 and the C-terminus will bind to the B-pocket and F-pocket of the HLA molecule, respectively[42]. Thus, each HLA molecule binds a large, yet finite, number of correctly processed peptides, with a well-defined recognition motif that is defined by the biochemical properties of the HLA binding pockets.

1.4.3. Recognition of peptide/HLA by the T-cell receptor (TCR)

CTL migrate throughout the body, scanning the surface of host cells and tissues for evidence of pathogen infection. CTL and antigen-presenting cell (APC) form an organized structure containing receptors mediating antigen recognition and adhesion, commonly referred to as the immunological synapse[43]. This structure allows the T-cell

receptor (TCR) present on the CTL to sample different pHLA ligands presented on the surface of the APC and to quickly initiate intracellular signalling events.

The TCR determines the antigen specificity of the CTL. It is a heterodimer, composed of an alpha and a beta chain, which are generated through recombination of Variable (V) and Joining (J) genes for the alpha chain, or V, Diversity (D), and J genes for the beta chain. The variable complementarity-determining regions (CDR) on the α and β subunits determine the pHLA specificity of the TCR. In general, the most highly variable CDR3, which is generated by the V(D)J recombination event, contacts the peptide directly, while the CDR1 and CDR2, which are encoded by the V gene sequence, interacts with the HLA molecule[44].

The CD8 co-receptor localizes with TCR on the T-cell surface and contributes to pHLA binding and signalling[45]. CD8 functions as a dimer comprised of two CD8α proteins or one CD8α and one CD8β protein. CD8 binds to a conserved motif on HLA class I proteins and stabilizes the TCR interaction with its pHLA ligand. The cytoplasmic tail of CD8 recruits p56-Lck tyrosine kinase to the immunological synapse, which enhances proximal signalling events following specific TCR/pHLA interaction through phosphorylation of CD3ζ (Figure 1.2). The resulting signal cascade leads to the activation of transcription factors, including the nuclear factor of activated cells (NF-AT). Expression of cellular genes initiates various CTL effector functions, such as target T-cell cytolysis, cytokine secretion, and clonal expansion.

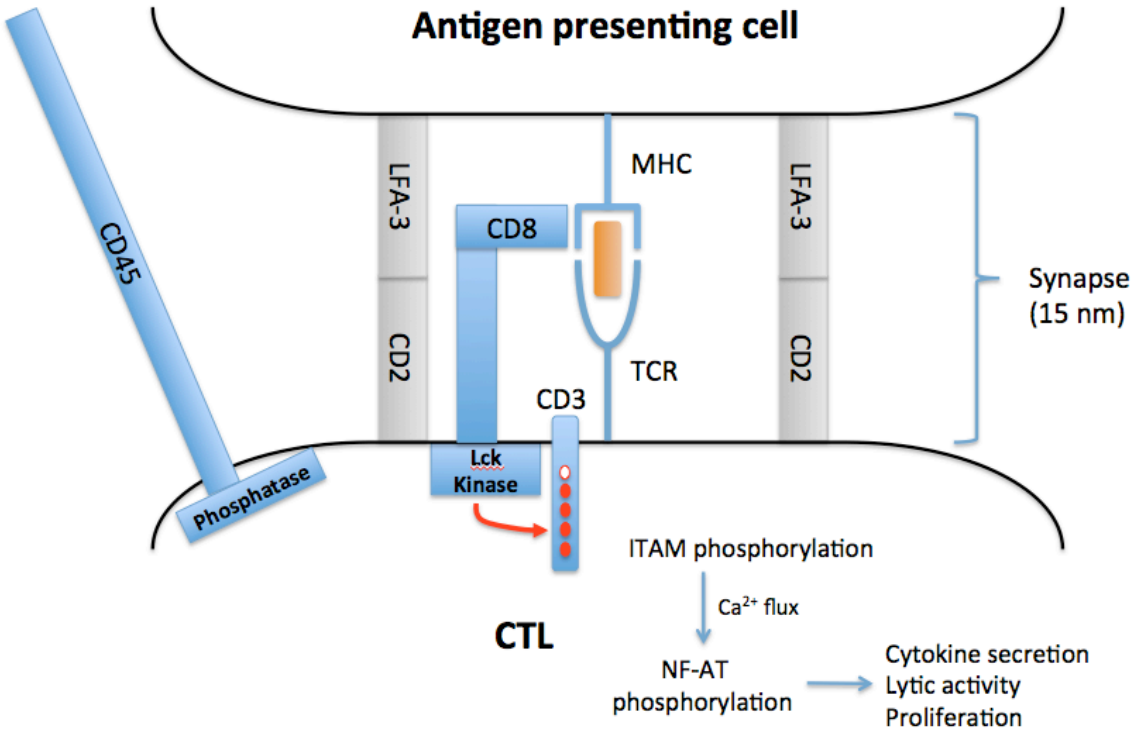

Figure 1.2 The immunological synapse.

T-cell activation begins with the formation of an immunological synapse. Specific TCR-pHLA interaction initiates a signalling cascade from the intracellular domain of the CD3 complex, which stimulates gene expression required for T-cell effector functions.

1.5. HLA-associated polymorphisms: viral sites under CTL pressure

HIV evades CTL immune responses through selection of viral mutations within or adjacent to targeted epitopes[46]. Because viral peptides are presented to CTL in complex with host HLA, and because HLA are genetically determined within a given individual, CTL-mediated immune selection pressure on the virus can be considered as HLA-dependent. Therefore, studying host HLA-associated polymorphisms in HIV sequences can help us to understand how the virus adapts in response to host CTL pressure, and may also provide a novel way to identify the *in vivo* characteristics of an effective antiviral CTL response. The application of statistical techniques[47] to large population-level datasets of linked host HLA and viral genotype has allowed for the systematic identification of HLA-associated immune escape pathways[37,46]. In brief, these methods take into account the effect of viral phylogeny when using cross-sectional viral

sequence data[48] and false positives that can arise due to HIV codon-codon covariation and host HLA linkage disequilibrium. The false discovery rate method[49] is used to control for multiple comparisons.

HLA-associated polymorphisms presumably represent immune escape variants (Figure 1.3). Indeed, numerous HLA-associated polymorphisms have been experimentally confirmed to compromise peptide-HLA binding[50], antigen processing[51], T-cell recognition of the pHLA complex[52], and killer immunoglobulin-like receptor (KIR) binding[53]. Furthermore, HLA-associated polymorphisms identified at the population level represent markers of viral sites under immune pressure by individual HLA alleles. Therefore, greater characterization of these polymorphisms may provide insight into the mechanisms of immune evasion and correlates of CTL-mediated protection.

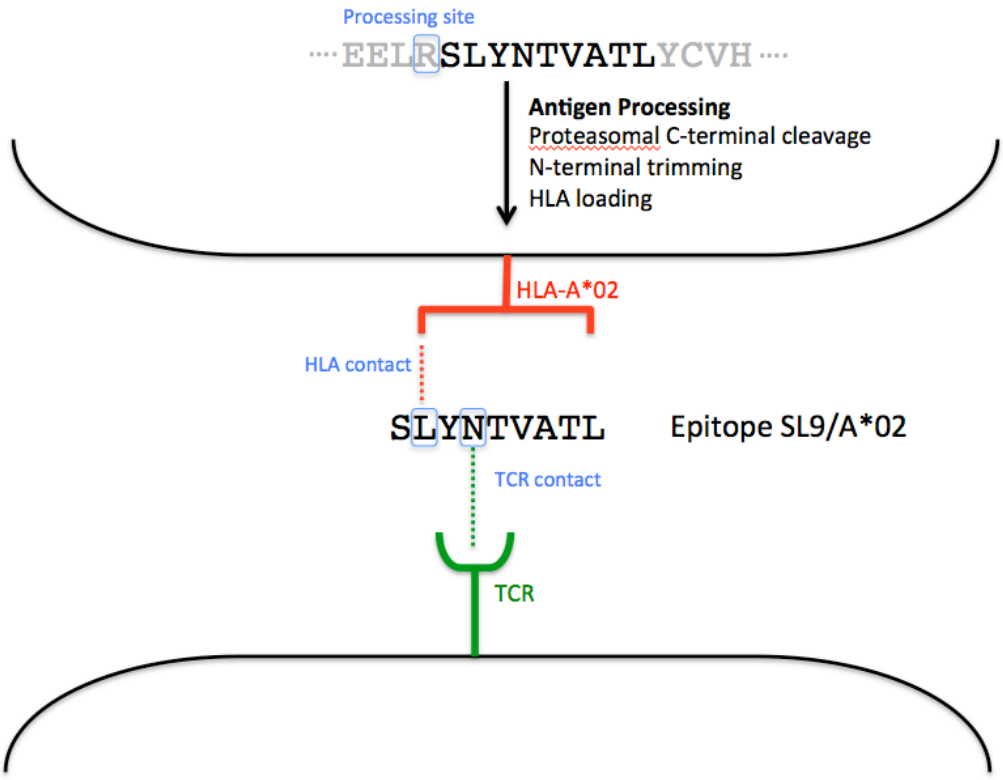

Figure 1.3 HLA-associated polymorphisms can disrupt CTL recognition.

The HIV-Gag SL9 epitope is shown as an example. Blue boxes denote residues, which can mediate CTL escape through different mechanisms. These mechanisms include the disruption of antigen processing, HLA/peptide binding, and TCR/pMHC binding. The TCR contact at position 4 is hypothetical.

1.6. Identifying correlates of CTL control is critical for vaccine design

Although there is substantial evidence that CTL are crucial immune determinants of clinical outcome, qualities of an effective CTL response – i.e. correlates of immune control of HIV infection or disease progression – have not been clearly established[54]. Identifying correlates of CTL immune control is a crucial step in the process to develop an HIV vaccine, since a vaccine will most likely need to induce T-cell responses with appropriate antiviral qualities to be effective. It is difficult to assess all potential correlates of immune control and to discuss them as discrete categories. However, in the case of HIV at least two factors must be considered: first, aspects of the virus, including mutational constraints that may affect sequence evolution and pathways of escape; and second, aspects of CTL themselves, such as effector function.

HIV mutational constraints could be a parameter of interest when designing a CTL vaccine. CTL responses targeting the more highly conserved Gag protein are associated with reduced viremia[55,56], indicating that immune recognition of more constrained regions may contribute significantly to relative viremia control and slower disease progression. In some cases, this appears to be due to the limited ability of HIV to evade CTL pressure through the development of escape mutations (i.e. B*27-KK10[57]), while in other cases, this may be due to impaired viral fitness resulting from the accumulation of escape mutations in important functional domains (i.e. B*57-TW10[58]). Computational analyses of the protective HLA allele B*57 suggested that a more diverse CTL repertoire could be available in B*57-expressing people[59]. A larger number of CTL responses would increase the chance that a given viral region is targeted; if these epitopes happen to be in highly conserved regions, placing immune pressure on them may result on the selection of escape mutations that confer substantial fitness cost. Similarly, a diverse CTL repertoire could mount several CTL responses within an epitope, reducing the number of viable escape pathways that have a minimal fitness consequence. Thus, CTL responses targeting more conserved regions of the virus, which select for escape mutations that results in reduced viral fitness, may represent a key immune correlate of HIV control.

Specific CTL effector functions, determined in large part by biochemical qualities of antigen-specific TCR recognition and signalling, may also be linked to viral control.

CTL clones that are able to detect lower levels of pHLA ligand on target T-cells, referred to as high functional avidity or high antigen sensitivity clones, suppress *in vitro* HIV replication better than lower avidity CTL[60]. Therefore, under conditions of low pHLA density, CTL with higher antigen sensitivity may be necessary to recognize infected cells. CTL responses restricted by HLA-B alleles, which tend to be more protective, display 10-fold higher functional avidity than clones restricted by HLA-A or HLA-C[61]. The HIV accessory protein Nef impairs CTL recognition by downregulating surface pHLA[62]; and indeed, recent evidence suggests that HLA-B tends to be associated with protection at least in part because it resists Nef-mediated downregulation to an extent[63]. Besides functional avidity, the binding motif of a TCR may influence the ability of virus to escape: TCR clonotypes that cross-recognize both the wild type epitope and a potential escape variant could be more protective. Functional avidity and cross reactivity may be inter-related phenotypes, as both are enriched in elite controllers[64]. The effect of TCR binding properties remains an underexplored factor that may reveal why individuals with the same HLA experience different clinical outcomes.

1.7. TCR binding properties

1.7.1. Functional avidity: a critical correlate of immune control?

Functional avidity, or *antigen sensitivity*, is the relative amount of peptide needed to induce a functional T-cell response[65]. It may be a key parameter that characterizes a highly protective CTL response. For T-cells, functional avidity is mediated largely by the biochemical affinity of the TCR/pHLA complex; however, it is also affected by variables in the T-cell, target T-cell, and by the cytokine environment (Figure 1.4). Surface pHLA density is determined by pathogen protein expression, antigen-processing efficiency, and HLA expression. Antigen processing proteins (TAP, ERAAP) and HLA expression are modulated by cytokines such as IFNg[66]. On the CTL, TCR and CD8 co-receptor expression are also variable [44,67]. CD8 expression can be modulated by common γ-chain cytokines (IL-2, IL-7, and IL-15) [68], and may affect functional avidity in low affinity TCR up to 1 million fold[69]. Co-inhibitory molecules such as PD-1 can block intracellular TCR signalling events by preventing Zap-70 phosphorylation[70]. Modulation of functional avidity has been theorized to be a *tuning mechanism* by Sewell and colleagues, through

which a T-cell can be made to be highly sensitive but promiscuous, or poorly sensitive but highly specific[44].

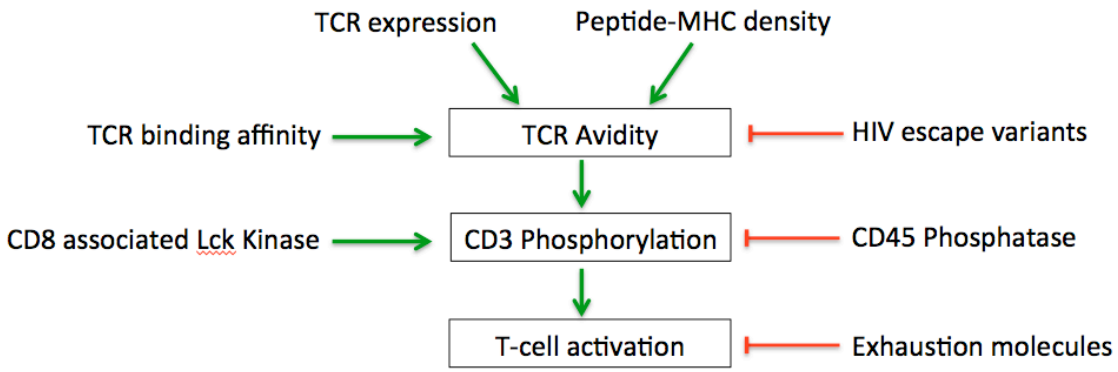

Figure 1.4 Variables that affect functional avidity.

Functional avidity is the consequence of a multi-step signalling cascade. Each step can be regulated by positive and negative factors. Variables such as patient genetics, their clinical status, and how long they have been infected could affect functional avidity. For example, patients with a secondary viral infection at the time of sample collection could have an inflammatory cytokine environment, which would increase the contribution of TCR/CD8 expression and peptide-MHC density.

1.7.2. Functional avidity is difficult to study ex vivo

The highly interconnected nature of T-cell signalling mechanisms makes *ex vivo* studies of primary cells difficult. Although assaying antiviral CTL *ex vivo* reveals a correlation between their functional avidity and HIV clinical outcome, there is limited opportunity to define the variables that determine differences in functional avidity.

1.7.3. TCR have cross-reactive specificity

TCR binding for pMHC ligands is highly degenerate. Sequencing approaches place a lower limit of $2.5*10^7$ different TCR in the total naive α/β CTL repertoire[71]. However, considering the potential to generate 10^{13} distinct 10-mer epitopes and assuming that each TCR is specific for one ligand, approximately 500,000 epitopes would need to be encountered before one was recognized. Theoretical considerations taking into account the number of distinct TCR, the number of CTL precursors necessary for a rapid immune response, and the number of possible epitopes, have predicted that one TCR may recognize as many as one million unique peptide ligands[72]. Sewell and colleagues the function of a CTL clone against a combinatorial peptide library and

identified epitope positions 4, 5, and 6 as key TCR contact points, with the TCR in their study displaying over a million possible ligands due to variability outside of this region[73].

1.7.4. The impact of TCR binding degeneracy on vaccine design

TCR binding degeneracy has consequences for HIV vaccine design. First, synthetic ligands with optimized HLA and TCR binding properties might be designed that are more immunogenic than the natural epitope (Figure 1.5a). In this manner, mouse studies have shown that immunization using optimized affinity peptides (with adjuvant) provided better control of colon tumours compared to the natural epitope[74]. Second, certain TCR clonotypes may be inherently more effective than others. A vaccine antigen might be designed that could selectively expand these clonotypes from the naïve pool (Figure 1.5b). Combinatorial peptide libraries have been used to isolate high affinity peptides binding a particular CTL clone to the exclusion of other CTL clones with the same epitope specificity[75]. In particular, selectively activating CTL clones that recognize both the wild type HIV epitope and a predicted escape variant may lead to enhanced CTL-mediated control of infection. Therefore, studying the binding motif of CTL may augment a vaccine design strategy designed to elicit particular CTL repertoires.

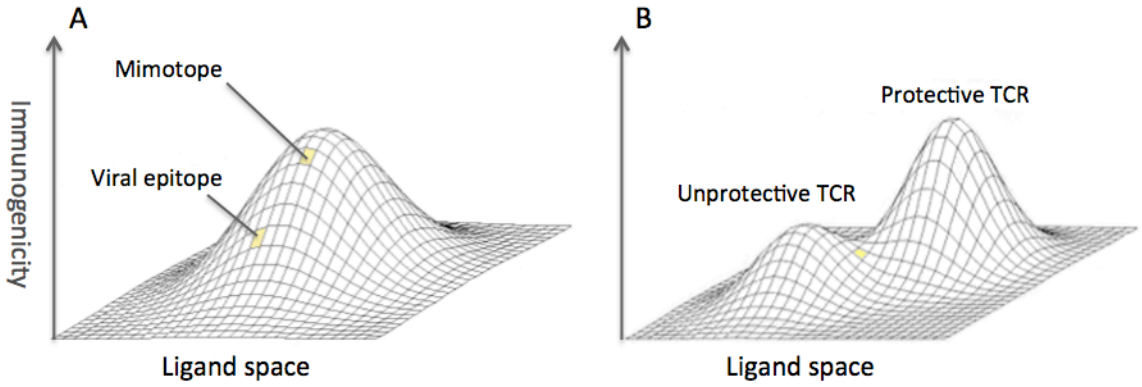

Figure 1.5 TCR bind a spectrum of ligands.

TCR bind a spectrum of ligands, and this can be used to guide vaccination strategies. a) A synthetic epitope (mimotope) may have a greater immunogenicity than the natural pathogen epitope. b) Two TCR can recognize the same epitope, but only one TCR might be protective. An immunogen could be chosen to specifically induce this protective TCR clonotype.

1.7.5. Cell-free assays can elucidate TCR/pHLA binding chemistry

TCR function can be studied using cell-free approaches, where TCR/pHLA binding is not affected by cellular variability. Surface plasmon resonance (SPR) and X-ray crystallography represent two such approaches. SPR has been used to study receptor / ligand binding without a tracer, and can provide direct estimates of K_{on}, K_{off}, and K_d. In brief, SPR involves the immobilization of TCR on a film mixed with soluble pHLA[76]. Using this approach, TCR binding to *monomeric* peptide-HLA has been measured to have weak, micromolar affinities[77]. X-ray crystallography is a powerful method for elucidating structural details of TCR-pHLA binding. Recently, a complete structure of TCR, pHLA-II, and CD4 has been published[78]. Measuring parameters such as K_d is appealing, as it is a constant value that does not vary between experiments. However, cell-free assays do not observe binding in the context of a cellular membrane, and do not take into account the various biological molecules (such as co-receptors and adhesion molecules), that can assist or inhibit this interaction.

1.8. Bridging TCR binding chemistry with immunological outcomes

1.8.1. The mechanism of TCR signalling has not been resolved

The mechanics surrounding the initial stages of TCR signal transduction remain unclear[79]. The low frequency of specific pHLA makes it difficult to envision a simple dimerization model of receptor signalling. Furthermore, the low affinity of TCR-pHLA binding is surprising given that a single specific pHLA molecule on a target T-cell can induce CTL killing[80]. However, the formation of an immunological synapse yields a structure that is enriched for multiple TCR molecules[81]. Several kinetic models of TCR – pHLA binding have been proposed to reconcile the results of these studies, including TCR segregation and serial triggering.

In the TCR segregation/redistribution model, the TCR does not undergo conformational changes (Figure 1.6a). Outside the synapse, TCR/CD3 ITAMs stochastically encounter kinases (p56-Lck) and phosphatases (CD45) at a basal rate, and no net phosphorylation occurs. However, within an immunological synapse, ITAM phosphorylation is enhanced: adhesion molecules mediate the formation of a narrow

close contact zone, and the larger CD45 phosphatase is prevented from approaching the TCR/CD3 complex. If the TCR does not bind pHLA, it drifts out of the contact zone, and reverts to the ground state of phosphorylation. However, specific TCR-pHLA binding causes a persistence of TCR/CD3 within the contact zone, which becomes increasingly phosphorylated[82]. In this model, no receptor dimerization is necessary and the spatial properties of the synapse are meaningful. An alternative model, the serial triggering model attempts to reconcile the observation that despite the low affinity of TCR/pHLA binding, several hours of signalling is necessary for T-cell activation. In this model, TCR bind and dissociate from specific pHLA repeatedly (Figure 1.6b). Thus, a TCR/pHLA interaction must have sufficient affinity to result in an interaction, but a low enough binding half life to detach and allow subsequent TCR to bind[83]. Thus, this model provides a satisfactory answer to how low affinity interactions can yield robust TCR signalling.

1.8.2. Cellular variables may explain conflicting results from biochemical studies

A better understanding of TCR recognition and signalling could lead to a more rational immunogen design process, but this has been hampered by paradoxical observations in the literature. While some studies have associated high affinity with optimal TCR activation[84], others suggest a high affinity can *inhibit* activation, with optimal ligands requiring an intermediate half life of interaction[85]. Indeed, it has been shown that T-cell signalling is more dependent on k_{off} at lower pHLA densities, while affinity is a better predictor of T-cell activation at higher pHLA densities[86]. Consequently, both binding affinity and the off-rate appear to be important parameters that affect TCR signalling, but these cannot be considered independently without taking a target T-cell variable, antigen density, into account.

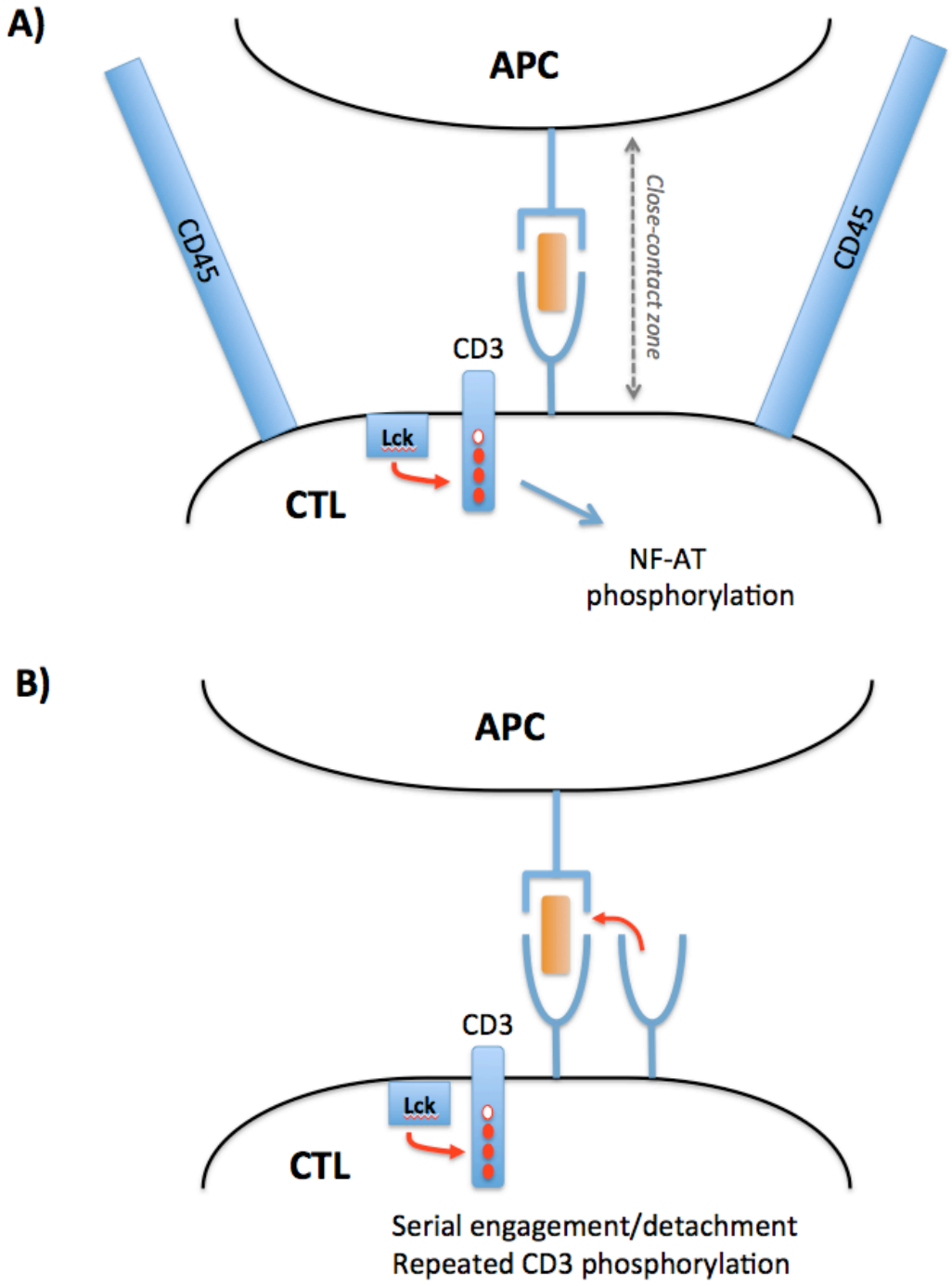

Figure 1.6 Proposed TCR signalling molecules.

A) Under the TCR redistribution/segregation model, the geometry of the immunological synapse causes a segregation of CD45 phosphatase from the TCR complex, causing a net phosphorylation of CD3 ITAMs. B) Under the serial engagement model, multiple TCR transiently sample the same pHLA repeatedly, and the additive effect of multiple TCR signal transduction events yields a productive T-cell activation signal.

1.8.3. Current measurements of functional avidity have limitations

Different T-cell-based approaches to study *ex vivo* functional avidity of CTL have been developed, each with certain advantages and limitations. If available, PBMCs can be assayed using high-throughput ELISPOT methods[87], with large numbers of different epitopes screened using peptide pools or libraries. CTL clones have been assayed using similar approaches, and the single specificity and sensitivity of CTL clones is appealing[88]. However, generating and maintaining a large number of unique CTL clones is challenging.

Cytokine secretion, as measured by ELISPOT, is a common approach to measure CTL functional avidity. Target T-cells are incubated with serial dilutions of peptide, and the amount of peptide yielding half-maximal signal (50% effector concentration; EC50) is recorded as a measure of functional avidity. ELISPOT is appealing because it is high throughput and sensitive to detect the total number of responding T-cells. However, ELISPOT assays are not designed to measure spot-size, which relates to the amount of cytokine secreted, and therefore it is difficult to determine a maximal cytokine intensity (E_{max}) value in this manner. Criticisms of potential bias in using ELISPOT have also been raised, as not all T-cell subsets secrete a given cytokine, such as IFN-gamma[89].

Peptide-HLA (pHLA) tetramer decay experiments can also be performed, using flow cytometry to measure the half-life for a given tetramer to dissociate from saturated T-cells. This assay is used to complement affinity measurements and takes multimeric TCR-pHLA binding into account[90]. Unfortunately, custom synthesis of pHLA tetramers is expensive and CD8-dependence of the TCR can affect studies of this type[91].

These methods may provide a more direct identification of the immune correlates of HIV control, since effective antiviral immunity may be related to the quality of an immune response, rather than a biophysical parameter of the TCR, *per se*. However, the interpretation of cell-based readouts from primary cells is difficult because it may be unclear which variables are driving the observed differences. Given that non-TCR variables can affect functional avidity, it becomes important to address the impact of both TCR and non-TCR factors on CTL activity.

1.8.4. A cell-line approach would allow for the controlled measurement of functional avidity

This thesis presents new methodology for studying functional avidity, using effector T-cell lines expressing the TCR of interest, and artificial antigen presenting target T-cells expressing a specific HLA allele restricting the epitope of interest. Although the details are presented later, we propose that this single platform can be used to study the effect of both TCR-specific variables, such as affinity and cross reactivity, and non-TCR variables, such as exhaustion markers and co-receptors. Because the effector and target T-cells are homogenous immortalized lines, much of the variability normally associated with *ex vivo* cell-based assays is reduced. This technique is sufficiently flexible to measure both EC50 and E_{max}, and the modular nature of the reporter allows different functions to be measured. Lastly, this assay has the mono-specificity normally only associated with using CTL clones, while remaining affordable and high throughput.

1.9. References

1. Barre-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science.* May 20 1983;220(4599):868-871.

2. Hecht R, Alban A, Taylor K, Post S, Andersen NB, Schwarz R. Putting it together: AIDS and the millennium development goals. *PLoS medicine.* Nov 2006;3(11):e455.

3. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry.* 1998;67:1-25.

4. Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science.* Jun 28 1996;272(5270):1955-1958.

5. Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature.* Jun 20 1996;381(6584):661-666.

6. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science.* May 10 1996;272(5263):872-877.

7. Kirchhoff F. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell host & microbe.* Jul 22 2010;8(1):55-67.

8. Petersen JL, Morris CR, Solheim JC. Virus evasion of MHC class I molecule presentation. *J Immunol.* Nov 1 2003;171(9):4473-4478.

9. Fauci AS. The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science.* Feb 5 1988;239(4840):617-622.

10. Kilmarx PH. Global epidemiology of HIV. *Current opinion in HIV and AIDS.* Jul 2009;4(4):240-246.

11. HIV/AIDS Epi Update 2010. In: Canada PHAo, ed2010.

12. Wood E, Montaner JS, Li K, et al. Burden of HIV infection among aboriginal injection drug users in Vancouver, British Columbia. *American journal of public health.* Mar 2008;98(3):515-519.

13. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *The New England journal of medicine.* Sep 11 1997;337(11):725-733.

14. Hogg RS, O'Shaughnessy MV, Gataric N, et al. Decline in deaths from AIDS due to new antiretrovirals. *Lancet.* May 3 1997;349(9061):1294.

15. Li TS, Tubiana R, Katlama C, Calvez V, Ait Mohand H, Autran B. Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. *Lancet.* Jun 6 1998;351(9117):1682-1686.

16. Butler DM, Smith DM, Cachay ER, et al. Herpes simplex virus 2 serostatus and viral loads of HIV-1 in blood and semen as risk factors for HIV transmission among men who have sex with men. *AIDS.* Aug 20 2008;22(13):1667-1671.

17. Das M, Chu PL, Santos GM, et al. Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco. *PloS one.* 2010;5(6):e11068.

18. Finzi D, Hermankova M, Pierson T, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science.* Nov 14 1997;278(5341):1295-1300.

19. Kaufmann DE, Lichterfeld M, Altfeld M, et al. Limited durability of viral control following treated acute HIV infection. *PLoS medicine.* Nov 2004;1(2):e36.

20. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science.* Nov 25 1988;242(4882):1171-1173.

21. Clavel F, Hance AJ. HIV drug resistance. *The New England journal of medicine.* Mar 4 2004;350(10):1023-1035.

22. Ciurea A, Klenerman P, Hunziker L, et al. Viral persistence in vivo through selection of neutralizing antibody-escape variants. *Proceedings of the National Academy of Sciences of the United States of America.* Mar 14 2000;97(6):2749-2754.

23. Leslie AJ, Pfafferott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine.* Mar 2004;10(3):282-289.

24. Izzedine H, Harris M, Perazella MA. The nephrotoxic effects of HAART. *Nature reviews. Nephrology.* Oct 2009;5(10):563-573.

25. Wilkins C, Gale M, Jr. Recognition of viruses by cytoplasmic sensors. *Current opinion in immunology.* Feb 2010;22(1):41-47.

26. Medzhitov R, Janeway CA, Jr. Innate immunity: impact on the adaptive immune response. *Current opinion in immunology.* Feb 1997;9(1):4-9.

27. Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations (*). *Annual review of immunology.* 2010;28:445-489.

28. Sheppard HW, Lang W, Ascher MS, Vittinghoff E, Winkelstein W. The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels. *AIDS.* Sep 1993;7(9):1159-1166.

29. Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity.* Sep 2007;27(3):406-416.

30. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science.* Mar 12 1999;283(5408):1748-1752.

31. Pereyra F, Jia X, McLaren PJ, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* Dec 10 2010;330(6010):1551-1557.

32. O'Brien SJ, Gao X, Carrington M. HLA and AIDS: a cautionary tale. *Trends in molecular medicine.* Sep 2001;7(9):379-381.

33. Feinberg MB. Changing the natural history of HIV disease. *Lancet.* Jul 27 1996;348(9022):239-246.

34. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of virology.* Jul 1994;68(7):4650-4655.

35. Hendel H, Caillat-Zucman S, Lebuanec H, et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol.* Jun 1 1999;162(11):6942-6946.

36. Schmitz JE, Kuroda MJ, Santra S, et al. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science.* Feb 5 1999;283(5403):857-860.

37. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science.* May 24 2002;296(5572):1439-1443.

38. Cascio P, Hilton C, Kisselev AF, Rock KL, Goldberg AL. 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *The EMBO journal.* May 15 2001;20(10):2357-2366.

39. Hearn A, York IA, Bishop C, Rock KL. Characterizing the specificity and cooperation of aminopeptidases in the cytosol and endoplasmic reticulum during MHC class I antigen presentation. *J Immunol.* May 1 2010;184(9):4725-4732.

40. Heemels MT, Ploegh HL. Substrate specificity of allelic variants of the TAP peptide transporter. *Immunity.* Dec 1994;1(9):775-784.

41. Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature reviews. Immunology.* Dec 2011;11(12):823-836.

42. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC immunology.* 2008;9:1.

43. Dustin ML, Groves JT. Receptor signaling clusters in the immune synapse. *Annual review of biophysics.* 2012;41:543-556.

44. Wooldridge L, Lissina A, Cole DK, van den Berg HA, Price DA, Sewell AK. Tricks with tetramers: how to get the most from multimeric peptide-MHC. *Immunology.* Feb 2009;126(2):147-164.

45. Laugel B, Cole DK, Clement M, Wooldridge L, Price DA, Sewell AK. The multiple roles of the CD8 coreceptor in T cell biology: opportunities for the selective modulation of self-reactive cytotoxic T cells. *Journal of leukocyte biology.* Dec 2011;90(6):1089-1099.

46. Brumme ZL, John M, Carlson JM, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PloS one.* 2009;4(8):e6687.

47. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology.* Nov 2008;4(11):e1000225.

48. Bhattacharya T, Daniels M, Heckerman D, et al. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science.* Mar 16 2007;315(5818):1583-1586.

49. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America.* Aug 5 2003;100(16):9440-9445.

50. Ammaranond P, Zaunders J, Satchell C, van Bockel D, Cooper DA, Kelleher AD. A new variant cytotoxic T lymphocyte escape mutation in HLA-B27-positive individuals infected with HIV type 1. *AIDS research and human retroviruses.* May 2005;21(5):395-397.

51. Draenert R, Le Gall S, Pfafferott KJ, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *The Journal of experimental medicine.* Apr 5 2004;199(7):905-915.

52. Iglesias MC, Almeida JR, Fastenackels S, et al. Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood.* Aug 25 2011;118(8):2138-2149.

53. Brackenridge S, Evans EJ, Toebes M, et al. An early HIV mutation within an HLA-B*57-restricted T cell epitope abrogates binding to the killer inhibitory receptor 3DL1. *Journal of virology.* Jun 2011;85(11):5415-5422.

54. Makedonas G, Betts MR. Living in a house of cards: re-evaluating CD8+ T-cell immune correlates against HIV. *Immunological reviews.* Jan 2011;239(1):109-124.

55. Jia M, Hong K, Chen J, et al. Preferential CTL targeting of Gag is associated with relative viral control in long-term surviving HIV-1 infected former plasma donors from China. *Cell research.* May 2012;22(5):903-914.

56. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine.* Jan 2007;13(1):46-53.

57. Schneidewind A, Brockman MA, Yang R, et al. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *Journal of virology.* Nov 2007;81(22):12382-12393.

58. Brockman MA, Schneidewind A, Lahaie M, et al. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. *Journal of virology.* Nov 2007;81(22):12608-12618.

59. Kosmrlj A, Read EL, Qi Y, et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature.* May 20 2010;465(7296):350-354.

60. Almeida JR, Sauce D, Price DA, et al. Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV-suppressive activity. *Blood.* Jun 18 2009;113(25):6351-6360.

61. Berger CT, Frahm N, Price DA, et al. High-functional-avidity cytotoxic T lymphocyte responses to HLA-B-restricted Gag-derived epitopes associated with relative HIV control. *Journal of virology.* Sep 2011;85(18):9334-9345.

62. Adnan S, Balamurugan A, Trocha A, et al. Nef interference with HIV-1-specific CTL antiviral activity is epitope specific. *Blood.* Nov 15 2006;108(10):3414-3419.

63. Rajapaksa US, Li D, Peng YC, McMichael AJ, Dong T, Xu XN. HLA-B may be more protective against HIV-1 than HLA-A because it resists negative regulatory factor (Nef) mediated down-regulation. *Proceedings of the National Academy of Sciences of the United States of America.* Aug 14 2012;109(33):13353-13358.

64. Mothe B, Llano A, Ibarrondo J, et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PloS one.* 2012;7(1):e29717.

65. Slifka MK, Whitton JL. Functional avidity maturation of CD8(+) T cells without selection of higher affinity TCR. *Nature immunology.* Aug 2001;2(8):711-717.

66. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. *Journal of leukocyte biology.* Feb 2004;75(2):163-189.

67. Xiao Z, Mescher MF, Jameson SC. Detuning CD8 T cells: down-regulation of CD8 expression, tetramer binding, and response during CTL activation. *The Journal of experimental medicine.* Oct 29 2007;204(11):2667-2677.

68. Park JH, Adoro S, Lucas PJ, et al. 'Coreceptor tuning': cytokine signals transcriptionally tailor CD8 coreceptor expression to the self-specificity of the TCR. *Nature immunology.* Oct 2007;8(10):1049-1059.

69. Holler PD, Kranz DM. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. *Immunity.* Feb 2003;18(2):255-264.

70. Sheppard KA, Fitz LJ, Lee JM, et al. PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3zeta signalosome and downstream signaling to PKCtheta. *FEBS letters.* Sep 10 2004;574(1-3):37-41.

71. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. *Science.* Oct 29 1999;286(5441):958-961.

72. Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology today.* Sep 1998;19(9):395-404.

73. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. *The Journal of biological chemistry.* Jan 6 2012;287(2):1168-1177.

74. McMahan RH, McWilliams JA, Jordan KR, Dow SW, Wilson DB, Slansky JE. Relating TCR-peptide-MHC affinity to immunogenicity for the design of tumor vaccines. *The Journal of clinical investigation.* Sep 2006;116(9):2543-2551.

75. Ekeruche-Makinde J, Clement M, Cole DK, et al. T-Cell Receptor optimized peptide skewing of the T-cell repertoire can enhance antigen targeting. *The Journal of biological chemistry.* Sep 5 2012.

76. Margulies DH, Plaksin D, Khilko SN, Jelonek MT. Studying interactions involving the T-cell antigen receptor by surface plasmon resonance. *Current opinion in immunology.* Apr 1996;8(2):262-270.

77. Eisen HN, Sykulev Y, Tsomides TJ. Antigen-specific T-cell receptors and their reactions with complexes formed by peptides with major histocompatibility complex proteins. *Advances in protein chemistry.* 1996;49:1-56.

78. Yin Y, Wang XX, Mariuzza RA. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proceedings of the National Academy of Sciences of the United States of America.* Apr 3 2012;109(14):5405-5410.

79. Dustin ML, Depoil D. New insights into the T cell synapse from single molecule techniques. *Nature reviews. Immunology.* Oct 2011;11(10):672-684.

80. Sykulev Y, Joo M, Vturina I, Tsomides TJ, Eisen HN. Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity.* Jun 1996;4(6):565-571.

81. Monks CR, Freiberg BA, Kupfer H, Sciaky N, Kupfer A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature.* Sep 3 1998;395(6697):82-86.

82. Davis SJ, van der Merwe PA. The kinetic-segregation model: TCR triggering and beyond. *Nature immunology.* Aug 2006;7(8):803-809.

83. Valitutti S, Lanzavecchia A. Serial triggering of TCRs: a basis for the sensitivity and specificity of antigen recognition. *Immunology today.* Jun 1997;18(6):299-304.

84. Kitchen SG, Bennett M, Galic Z, et al. Engineering antigen-specific T cells from genetically modified human hematopoietic stem cells in immunodeficient mice. *PloS one.* 2009;4(12):e8208.

85. Corse E, Gottschalk RA, Krogsgaard M, Allison JP. Attenuated T cell responses to a high-potency ligand in vivo. *PLoS biology.* 2010;8(9).

86. Gonzalez PA, Carreno LJ, Coombs D, et al. T cell receptor binding kinetics required for T cell activation depend on the density of cognate ligand on the antigen-presenting cell. *Proceedings of the National Academy of Sciences of the United States of America.* Mar 29 2005;102(13):4824-4829.

87. Almeida JR, Price DA, Papagno L, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine.* Oct 1 2007;204(10):2473-2485.

88. Chen H, Ndhlovu ZM, Liu D, et al. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nature immunology.* Jul 2012;13(7):691-700.

89. Betts MR, Nason MC, West SM, et al. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood.* Jun 15 2006;107(12):4781-4789.

90. Holmberg K, Mariathasan S, Ohteki T, Ohashi PS, Gascoigne NR. TCR binding kinetics measured with MHC class I tetramers reveal a positive selecting peptide with relatively high affinity for TCR. *J Immunol.* Sep 1 2003;171(5):2427-2434.

91. Walker LJ, Sewell AK, Klenerman P. T cell sensitivity and the outcome of viral infection. *Clinical and experimental immunology.* Mar 2010;159(3):245-255.

2. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1

Published as: Carlson JM*, Brumme CJ*, **Martin E*,** Listgarten J, Brockman MA, Le AQ, Chui CK, Cotton LA, Knapp DJ, Riddler SA, Haubrich R, Nelson G, Pfeifer N, Deziel CE, Heckerman D, Apps R, Carrington M, Mallal S, Harrigan PR, John M, Brumme ZL; the International HIV Adaptation Collaborative. *Journal of Virology*. 2012 86(24):13202-116.

* indicates equal contribution

Supplementary tables S1, S2, and S3 and figures S1, S2, S3, and S4 are available online[1]. My contributions to this study included preparing the dataset of patient HLA and HIV data from which HLA-associated polymorphisms were identified, analyzing the position distribution of polymorphisms within epitopes, and estimating the effect of polymorphisms on HLA binding. I contributed most to results 2.3.3 and 2.3.4.

2.1. Introduction

HIV-1 is notorious for its genetic diversity and its ability to adapt to selection pressures[2-4]. Despite this, within-host HIV-1 evolution in response to antiretroviral [5,6], host cellular immune[7-11], antibody[12] and vaccine-induced[13] selection pressures occurs along generally predictable mutational pathways[14,15]. Studying these evolutionary pathways can offer insight into the immunopathogenesis of HIV-1 and may help inform the design of immune-based interventions and vaccines.

Substantial progress has been made in our understanding of HIV-1's ability to evade HLA class I restricted CD8+ CTL. In particular, application of novel statistical methods[14,16,17] to large population-based datasets of linked host and viral genetic information have facilitated the systematic identification of HLA-associated immune escape and covarying mutations in HIV-1[18-23], revealing important insights into HIV-1 adaptation to its host. We now appreciate that immune selection represents a major

force shaping HIV-1 diversity [14,15,24] and that HIV-1 escape pathways are generally predictable in context of host HLA allele expression[14,15,18]. Immune escape mutations occur within and outside CTL epitopes[16,19,21-23], and can compromise peptide-HLA binding[25,26], disrupt intracellular antigen processing[27,28], affect T-cell recognition of the peptide-HLA complex[29-33] and/or potentially affect killer immunoglobulin-like receptor (KIR) binding[34,35]. While some escape pathways are likely to be universal across HLA alleles and/or HIV-1 subtypes, widespread examples of differential escape between infected populations[23,36] and between closely related HLA class I alleles[37-39] have also been demonstrated. Population-level studies have also allowed us to estimate rates[40] and clinical implications [20,41] of immune escape, infer fitness costs of specific mutations[20], discover novel epitopes in conventional[16,42] and cryptic[43,44] reading frames, and forecast the consequences of continued immune-mediated adaptation for the future of the epidemic[45,46]. These and other findings have led to recommendations that escape information be incorporated into HIV-1 vaccine strategies, for example via immunogens that incorporate variant sequences[16,47] and/or that are located in mutationally constrained regions where viral escape mutations would be anticipated to incur substantial fitness costs[48-50].

Although population-level studies have significantly advanced our understanding of the specific mutational pathways whereby HIV-1 escapes from HLA-restricted immune pressures, much remains to be learned. The key to further exploring these data lies in the recognition that HLA-associated polymorphisms identified at the population level serve as markers of viral sites under strong *in vivo* immune pressure by individual HLA alleles – sufficiently strong that the virus responds by escape. As such, analysis of the distribution, frequency, location, statistical strength, and sequence conservation of HLA-associated polymorphic sites identified at the population level offers a unique perspective from which to identify (albeit indirectly) specific features that render certain immune responses more effective at controlling HIV progression than others. In other words, systematic analysis of HLA-driven evolutionary “imprints” on the HIV proteome can help identify correlates of protective cellular immunity.

In the largest population-level, proteome-wide analysis of immune escape in HIV-1 undertaken to date, we refine existing immune escape data by reporting HLA-associated polymorphisms at HLA supertype, type and subtype-level resolutions across the viral proteome in an HIV-1 subtype B context. We analyze this well-powered dataset

to investigate the characteristics of HLA-driven immune selection on HIV-1 in order to draw inferences regarding the general mechanisms HIV-1 uses *in vivo* to evade such pressures. Furthermore, by treating escape mutation pathways as evolutionary markers of strong HLA-restricted immune pressure by CTL, we identify specific features of these responses that differentiate protective from non-protective HLA class I alleles.

2.2. Methods

2.2.1. International HIV adaptation collaborative (IHAC)

The International HIV Adaptation Collaborative (IHAC) is an open multicenter cohort of chronically infected antiretroviral-naïve individuals from Canada, the USA and Australia for whom HLA class I and nearly full genome HIV plasma RNA sequences have been characterized. The present study was restricted to 1888 HIV-1 subtype B infected individuals (>95% of total cohort), including 1103 individuals from the British Columbia HOMER cohort (British Columbia, Canada)[18,22], 247 individuals from the Western Australian HIV Cohort Study (WAHCS, Western Australia, Australia)[14,16,51], and 538 US AIDS Clinical Trials Group (ACTG) protocol 5142 participants[23,52] who also provided human DNA under ACTG protocol 5128[53]. This newly expanded cohort is approximately one-third larger than that studied in 2009[22], and now features full-proteome HIV coverage. Ethical approval was obtained from Providence Health Care/University of British Columbia (HOMER cohort); Royal Perth Hospital Ethics Committee (WAHCS); and the NIH's National Institute of Allergy and Infectious Diseases (NIAID) Clinical Science Review Committee (CSRC) (ACTG 5142/5128).

2.2.2. HIV-1 sequencing

Nearly full-genome plasma HIV-1 RNA sequencing (all regions except gp120) for the HOMER cohort was performed at the BC Centre for Excellence in HIV/AIDS (BC CfE) as described[22]. Briefly, HIV RNA was extracted from plasma using standard methods and regions of interest were amplified by nested RT-PCR using HIV-specific primers. Amplicons were bulk sequenced on an Applied Biosystems 3100, 3700 and/or 3730xl automated DNA sequencer. Data were analyzed using 'Sequencher' (Genecodes) or custom software RECall[54]. Nucleotide mixtures were called if secondary

peak *height* exceeded 25% of the dominant peak (Sequencher), or if secondary peak *area* exceeded 20% of the dominant peak (RECall). Nearly full-genome HIV-1 sequencing for the WAHCS and ACTG 5142/5128 cohort participants was performed at the Centre for Clinical Immunology and Biological Statistics (CCIBS) laboratory in Perth, Australia as described in [22,23]. Plasma HIV RNA was extracted using standard methods and nearly complete viral genomes amplified using nested RT-PCR. Amplicons were bulk-sequenced using an Applied Biosystems 3730xl automated sequencer. Data were analyzed using semi-automated ASSIGN software with a nucleotide mixture threshold of 15% after consideration of the signal/noise ratio, yielding near-full genome sequences. The BC CfE and CCIBS previously undertook blinded inter-laboratory genotyping quality control comparisons and observed excellent inter-site concordance ([22] and unpublished data).

Non subtype-B sequences were identified by comparison to subtype references in the Los Alamos HIV Database using Recombination Identification Program (RIP, available at http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html), and removed from the analysis. HIV-1 sequences were aligned to subtype B reference strain HXB2 (Genbank Accession No. K03455). Final HLA/HIV sequence dataset sizes were as follows: Gag (N=1548), Pol (N=1799) [PR/RT, N=1786, INT, N=1566], Nef (N=1685), Vif (N=1325), Vpr (N=1310), Vpu (N=1243), gp120 (N=655), gp41 (N=1425), Tat (N=1734), Rev (N=1731). HIV-1 sequences from the ACTG, Perth and HOMER cohorts were previously deposited in GenBank[15,18,22,23,41]. Accession numbers for additional HOMER sequences are JX147023 - JX147784 (gp41), JX147785 - JX148365 (Tat/Rev exon1 region), JX148366 - JX148914 (Vpu) and JX148915 - JX149509 (vif). Linked HLA/HIV datasets from the BC HOMER cohort are available for sharing with interested researchers in accordance with UBC/Providence Health Care REB protocols; please contact the corresponding author for information.

2.2.3. HLA class I sequence-based typing and subtype imputations

HLA class I typing for the HOMER cohort was performed at the BC CfE using an in-house sequence-based typing protocol and interpretation algorithm[18,22,55], yielding a mixture of intermediate- and high (subtype-level)- resolution data. High-resolution HLA class I typing for the WAHCS and ACTG 5142/5128 cohorts was performed at the CCIBS as described in [22,23]. Allele interpretation was performed using ASSIGN [23].

In previous studies we had addressed the issue of mixed-resolution HLA data by truncating all data to the two-digit (type) level[18,22,41]. However, in the present study we wished to identify HLA-associated polymorphisms at the supertype and subtype levels, which requires high-resolution data. Therefore, for all individuals with missing, low or intermediate level resolution at one or more loci (711 of 1888; 38%), we employed a machine learning algorithm trained on a dataset of complete high resolution HLA-A, B and C types from >13,000 individuals with known ethnicity ([56]; available at http://research.microsoft.com/en-us/projects/bio/mbt.aspx#HLA-Completion), to complete the data to high resolution at all loci. The resulting output lists the inferred possible HLA-A, B and C haplotypes, along with their respective probabilities, for each individual. Rather than assigning each individual their highest probability HLA haplotype, downstream analyses incorporate all HLA haplotypes with probabilities >1% as weighted averages for each individual. HLA types could not be imputed when data were missing from 2 (or more) loci. In such cases, an HLA haplotype containing null values for the missing data, and a probability of 1.0 was used.

2.2.4. Identification of HLA-associated polymorphisms

HLA-associated polymorphisms were identified using phylogenetically informed methods as previously published[17,22,39,57], with some modifications. Briefly, a maximum likelihood phylogenetic tree is constructed for each gene and a model of conditional adaptation is inferred for each observed amino acid at each codon. In this model, the amino acid is assumed to evolve independently along the phylogeny, until it reaches the observed hosts (tree tips). In each host, the selection pressure arising from HLA-mediated T-cell responses and amino acid covariation is directly modeled using a weighted logistic regression, in which the individual's HLA repertoire and covarying amino acids are used as predictors and the bias is determined by the transmitted sequence[39]. Because the transmitted sequence is not observed, we average over the possible transmitted sequences, as inferred from the phylogeny.

We extended this approach to incorporate predicted high resolution HLA data as follows. For each subject in the cohort, a number of "fractional" individuals are generated, representing each possible transmitted polymorphism as well as each possible completed HLA repertoire, with fractional weights representing the probability of observing the given HLA repertoire and transmitted polymorphism, as determined by the

HLA completion (see above) and phylogeny, respectively. When the observed polymorphism represents a mixture, we extend the weighting scheme to represent all possible mixture outcomes, with relative weights inversely proportional to the number of amino acids in the mixture. Maximum likelihood estimation for logistic regression and phylogenetic parameters is jointly estimated using expectation maximization[58].

To identify which factors contribute to the selection pressure, a forward selection procedure is employed, in which the most significant association is iteratively added to the model, with p-values computed using the likelihood ratio test. To increase our statistical power, each codon is divided into a set of binary variables, one for each observed amino acid. In addition, we only consider HLA alleles and amino acids that are observed in at least 10 individuals in the population, or for which at least 10 individuals do *not* express the HLA or polymorphism. In the case of imputed HLA alleles, we use the expected number of individuals expressing that allele, where expectation is taken with respect to the HLA completion probability. Statistical significance is reported using q-values, the p-value analogue of the false discovery rate (FDR), for each p-value threshold [59]. The FDR is the expected proportion of false positives among results deemed significant at a given threshold; for example, at a q≤0.2, we expect 20% of identified associations to be false positives. We compute q-values separately for each protein.

HLA-associated polymorphisms are grouped into two categories: (1) amino acids significantly enriched in the *presence* of the HLA allele in question ("adapted" forms), and (2) amino acids significantly enriched in the *absence* of the HLA allele in question ("nonadapted" forms). To provide an example, the B*27-associated R264K is a well-known escape mutation that commonly occurs at position 2 of the KK10 epitope in $p24^{Gag\ 25}$. At this codon, "R" represents the B*27-associated "nonadapted" form while "K" represent the B*27-associated "adapted" form. In general, "nonadapted" forms correspond to the subtype B consensus sequence while "adapted" forms correspond to polymorphic variants, but exceptions occur. In addition, we will sometimes differentiate between "Direct" and "Indirect" HLA-associated polymorphisms: the former represent associations that are detected when conditioning on HIV sequence covariation, and the latter arise only when covarying amino acids are not considered (and thus may include compensatory mutations). Analyses in this paper employ all observed Direct and Indirect associations.

2.2.5. Verifying the impact of HLA imputation on the ability to identify HLA-associated polymorphisms

As described above, some level of HLA imputation was required for 711 of 1888 (38%) of individuals in the present study. To investigate the impact of HLA imputation on identification of HLA-associated polymorphisms in a "worst-case" scenario, we took the subset of patients for whom full high-resolution data were available (N=1177), truncated all HLA types to two-digit resolution and re-imputed them to high resolution. The highest-probability imputed allele matched the original call in 93% of cases. We then identified HLA-associated polymorphisms from the imputed dataset and compared to those obtained from analysis of the same dataset at high-resolution (not shown). When matching over unique HLA (type)-HIV codon pairs, 82% of associations derived from the "imputed" dataset matched those derived from the high-resolution dataset, a result that is expected given the false-discovery rate of 0.2 and the expected loss of statistical power when analyzing imputed data.

2.2.6. Testing for differential escape

The statistical model for differential escape among HLA subtypes belonging to a given allele group (type) was defined as previously described[39]. However, whereas the previous study[39] explicitly investigated differential escape within published epitopes, here we tested for differential escape in general across the entire proteome: meaning, cases where one subtype selects a given polymorphism and another does not select for that polymorphism, or selects for that polymorphism at a lower frequency. As such, "differential escape" in the present context could be due to less effective (or lack of) responses to specific epitopes. Briefly, the phylogenetically corrected logistic regression model was used to test for evidence that the odds of selection differed between an HLA type and a particular subtype. For example, to test if the probability *p* of escape to 242N (in $p24^{Gag}$) differed between B*58:01 and the rest of the B*58 subtypes, we used the phylogenetically corrected logistic regression model defined by $\ln\left(\frac{p}{1-p}\right) = a(B58) + b(B5801) + cT$, where *T* is a -1/1 binary variable representing the transmitted polymorphism, B58 and B5801 are 0/1 binary variables indicating whether the individual has the HLA allele in question, and the parameters (*a*, *b*, *c*) are chosen to maximize the likelihood of the data. The likelihood ratio was then used to test the null

hypothesis that *b*=0 (*i.e.*, that knowing that an individual expresses B*58:01 confers no additional information beyond knowing that the individual expresses B*58).

To validate the distinction between associations identified at type versus subtype-level resolution, we applied the above model as follows. For each association originally identified at the subtype level, we tested for differential escape between the identified subtype and the corresponding HLA type. If the resulting q-value was less than 0.2, we considered the test an instance of true differential escape (differential escape was confirmed in >80% of cases where escape was originally defined at the subtype level). For each association originally defined at the type level, we tested for differential escape against all subtypes observed in at least 10 individuals in the subset for whom high-resolution HLA data were available across all loci (N=1177 of 1888, 62%). Together, a total of 29 HLA types with at least two subtype members were available for analysis. If the q-value of the most significant such test was less than 0.2, we considered the HLA-polymorphism pair to represent a true instance of differential escape (a re-classification that occurred in <20% of cases where escape was originally defined at the type level). For both positional analyses and analyses involving the stratification of tests by HLA type, we limited the analysis to the most significant HLA-type/HIV-position pair to avoid double counting nonadapted/adapted pairs.

2.2.7. Calculation of median genetic distances among HLA alleles of a given type

To estimate relative protein distances among HLA subtypes included in our differential escape analysis (see above), we retrieved their exon 2 and 3 amino acid sequences from *http://hla.alleles.org/data/txt/classi_prot.txt*[60] and used the *protdist* program from the PHYLIP software package[61] to calculate pairwise distances using the Henikoff Tillier Probability Matrix from Blocks (PBM) model of amino acid replacement[62]. For each HLA allele group (type), the median pairwise genetic distance between subtype members ("intratype distance") was calculated.

2.2.8. Definition of optimally-defined CD8+ epitopes

The Los Alamos optimal CD8+ epitope list comprises all published epitopes defined according to rigorous *in vitro* epitope fine-mapping and HLA restriction

experiments http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html (2009-08-31 update)[63] . However, optimal epitopes have not necessarily been tested in context of all possible restricting HLA alleles, nor are the alleles defined at the same resolution (*i.e.*: HLA type, subtype or serotype) throughout. To address these biases (and in recognition that alleles with shared properties are likely to present similar peptides[64]), we expanded each optimal epitope to include all members of the HLA type, supertype, or serotype to which the published restricting allele belonged, as described in [39]. Briefly, optimally defined epitopes were retrieved from the Los Alamos Database and hand-edited to include recently published epitopes ([38,65-71] and A. Bansal and P. Goepfert, personal communication). For optimal epitopes restricted by HLA alleles defined at type or subtype-level resolution, we expanded to include all subtypes belonging to the supertype to which the restricting allele belonged [64]. For example, an epitope originally defined as B*57:01-restricted was assigned to all members of the B58 supertype. For epitopes restricted by alleles defined at the serotype level (N=38), we expanded the list to include all HLA alleles belonging to that serotype[72], then identified supertypes common to a majority of the resulting expansion and expanded to include all subtypes matching those supertypes. If no supertype was defined (as is the case for HLA-C alleles), we expanded the list to include all HLA subtypes belonging to the HLA type of the original restricting allele. The full list of "expanded" epitopes (along with their original restrictions) is provided in Table S1.

2.2.9. Analysis of population-level immune escape pathways to identify correlates of protective immunity: data definitions

Protective effects of specific HLA class I alleles on HIV progression were defined according to published hazard ratios for progression to AIDS (HR-AIDS), determined for 54 individual class I alleles at type-level (2-digit) resolution in a natural history cohort of 600 Caucasian seroconverters[73]. To allow analysis at subtype-level resolution using HLA-associated polymorphisms specific to the population in which HR-AIDS were originally derived, we assigned each HR to the most frequently observed HLA subtype in Caucasians (e.g.: A*02 was assigned to A*02:01), and analyzed only HLA-associated polymorphisms restricted by these subtypes. In both univariate and multivariate analyses, HR-AIDS values were log-transformed to render their distribution more normal.

HLA-associated polymorphism-related variables that we investigated could be classified into two broad categories: features related to the frequency and/or distribution of sites under selection by a given HLA allele (defined as unique HIV codons harboring a polymorphism associated with that HLA allele), and features related to the signal strength of sites under HLA selection and/or to evolutionary constraints on the sites themselves. The first category contained the following variables: 1) Number (#) of sites under HLA selection (computed as the total number of HLA-associated polymorphic sites in HIV-1; both overall and within/near published epitopes only); 2) Proportion (%) of sites under HLA selection, by protein (computed as the proportion of the total HLA-associated sites occurring within the HIV-1 protein of interest); 3) Number (#) of sites under HLA selection occurring at anchor residues (computed as the total number of HLA-associated sites occurring at anchor positions within known epitopes); 4) Proportion (%) of sites under HLA selection occurring at anchor residues (proportion of total HLA-associated sites inside or within +/- 3aa of an epitope, that occurred at an anchor position); 5) Number (#) of "Active" epitopes (defined as the number of published epitopes harboring HLA-associated polymorphisms); and 6) Median number (#) of sites under HLA selection per active epitope. The second category contained the following variables: 7) Median odds ratio (OR) of selection (computed by taking the maximum absolute log odds ratio of all nonadapted and adapted associations per HLA at each unique site, followed by computing the median of all sites per HLA, where the odds ratio for a given HLA-polymorphism pair was computed as the ratio of the odds of observing the polymorphism among individuals expressing the allele to the odds of observing the polymorphism among individuals not expressing the allele); 8) Median conservation of sites under HLA selection (defined as the median sequence conservation of all sites associated with a particular HLA, where conservation was defined to be the proportion of individuals with the consensus sequence among individuals who did not express any HLA alleles associated with that site) [50]; and 9) Median number (#) of covarying codons per site (computed by summing the number of unique HIV-1 codons identified as covarying with each site, and computing the median for all sites per HLA). Where appropriate, analyses were undertaken at both the proteomewide and individual-protein levels (defined as Gag, Pol, Env, Nef and Accessory) to identify location-specific correlates of protection.

Multivariate analyses were attempted using a small number of distinct variables significant at the univariate level. To maximize the number of HLA alleles included in the multivariate analysis, alleles with "missing" variables (e.g.: those excluded from the original univariate analysis of OR of escape in Gag because they possessed no escape sites in that protein) were assigned the mean observed value in the dataset. Selected variables were regressed against log-transformed HR-AIDS values by automated forward selection using a stepwise Akaike Information Criterion (AIC) procedure [74].

2.3. Results

2.3.1. Identification of HLA-associated polymorphisms at three levels of HLA resolution

HLA-associated polymorphisms were identified across the entire HIV-1 proteome in an international cohort of N=1888 treatment-naïve, chronically subtype B-infected, individuals using phylogenetically informed methods that incorporated corrections for HIV codon covariation and HLA linkage disequilibrium [17,22,39] and used a q-value correction for multiple tests [59]. Missing, low or incomplete-resolution HLA data were imputed to subtype-level resolution using a published machine learning algorithm [56], and associations were computed by averaging over all possible HLA resolutions (see methods); this represents the first use of HLA completion algorithms in population-level immune escape analyses.

Studies of HLA-associated polymorphisms have generally featured individual pairwise analyses at the HLA type (*i.e.* 2-digit) and/or subtype (*i.e.* 4-digit) levels [17-19,22]. Supertype-level analyses have been uncommon even though statistical power could be enhanced to identify universal escape pathways within epitopes capable of binding related alleles. We therefore report HLA-associated polymorphisms at three different levels of resolution: first, at the combined supertype/type/subtype levels, second at the combined type/subtype levels, and third at the subtype level only. In the former two analyses, the level of resolution yielding the lowest p-value for each HLA-associated polymorphism is reported (that is, if the statistical signal for the association is stronger at the supertype compared to the type or subtype levels, then it is reported at the supertype level). A full list of HLA-associated polymorphisms identified at all three levels of

resolution are available online[1] in Table S2, while a full list of intraprotein codon covariation pathways is provided in Table S3. In addition, protein-specific "immune escape maps" for all associations q≤0.05 identified at the broadest (supertype/type/subtype) resolution level analysis are provided in Figure S1.

At q<0.2, over 2100 unique HLA-associated escape pathways (defined as those unique over HLA restriction, HIV codon co-ordinate, amino acid and direction of escape) occurring at over 750 HIV-1 codons were identified (Table S2). When the inherent sequence conservation of each HIV protein was taken into consideration, the proportion of sequence variation attributable, at least in part, to HLA selection pressures (defined as the proportion of sites harbouring at least one HLA-associated polymorphism) differed markedly by viral protein (Figure 2.1). For example, with 59% of its residues exhibiting ≥99.5% amino acid conservation, $p24^{Gag}$ is the most highly conserved HIV-1 protein, yet HLA pressures influence sequence variation at 40% of its variable codons. HLA-mediated polymorphisms were observed most frequently in the highly immunogenic Nef protein, where 67.4% of variable codons harbour HLA associations ($p=1.9x10^{-18}$; Fisher's exact test compared to other proteins). Conversely, the highly variable Vpu protein exhibited the least evidence of HLA-driven evolution (26.3% [p=0.009]; Figure 2.1) [18,23,75]. Across the entire HIV-1 proteome, polymorphisms at 804 of 1481 (35.2%) variable sites were associated with at least one HLA allele.

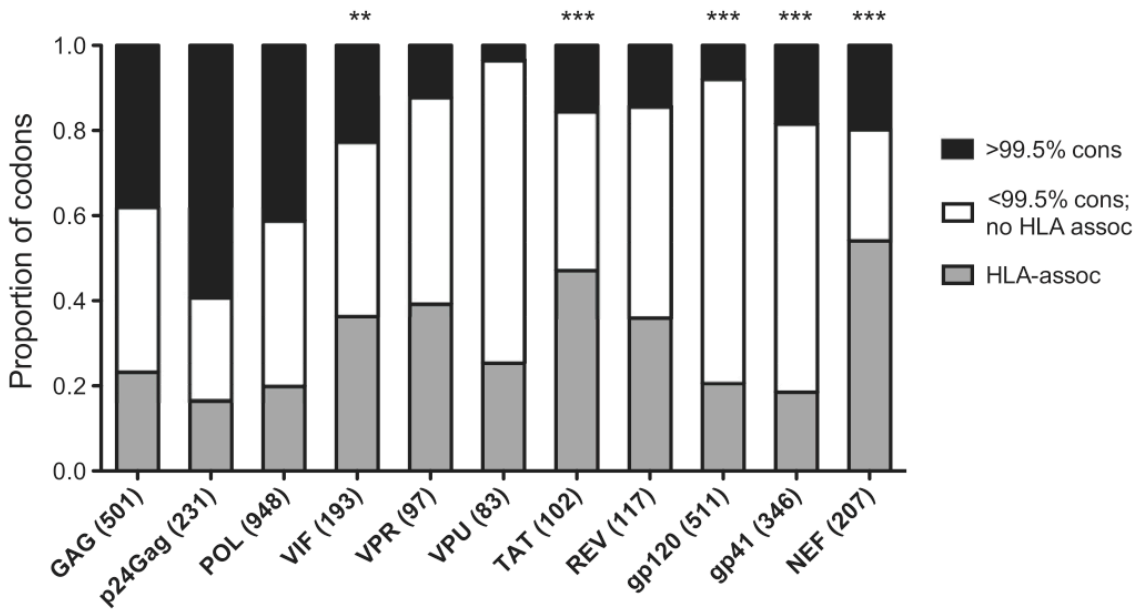

Figure 2.1 Sequence variation attribute to HLA-associated pressure.

The proportion of sequence variation attributable to HLA-associated selection pressure differs markedly by HIV-1 protein. The proportion of codons exhibiting >99.5% amino acid conservation at the population level in our cohort and the proportion of variable codons (those exhibiting < 99.5% amino acid conservation) harbouring no known HLA-associated polymorphisms are indicated, along with the proportion of codons harbouring at least one HLA-associated polymorphism. Numbers in parenthesis in the x-axis labels indicate the total length in amino acids of each protein (including the stop codon, if present). doi:10.1128/JVI.01998-12.

2.3.2. The extent of differential CTL escape between subtype members varies markedly by HLA

Although the HLA-associated polymorphisms identified at the three different levels of HLA resolution are highly concordant (Table S2 and data not shown), each analysis features relative merits and limitations. Analysis at broader resolution can enhance statistical signal when related alleles bind the same epitope and escape along the same pathway. For example, the Gag-TW10 epitope binds many members of the B58 supertype. At Gag codon 242 (position 3 of this epitope), relative odds of escape are 193:1 for B58 supertype members compared to non-B58 supertype members ($p=1.6x10^{-117}$). However, when relative odds are calculated at the type (e.g. B*57) or subtype (e.g. B*57:01) levels, lower odds are observed (128:1, $p=1.2x10^{-87}$ and 114:1, $p=5.4x10^{-72}$ respectively) because persons harbouring closely related alleles (e.g.: B*58 and/or B*58:01) are classified as lacking the specific allele under investigation, even

though they also select the same escape mutation. Conversely, when alleles belonging to the same type or supertype bind different peptides or select different escape pathways, analyzing at broader resolution may yield inappropriate results while still achieving high levels of statistical significance, and may mask differential escape (where similar HLA alleles bind the same epitope but escape along distinct pathways) [39]. For example, despite belonging to the B58 supertype, B*58:02 does not select for the T242N escape mutation (Table S2).

Notably, even when analyzed at the broadest resolution level, the majority of HLA-associated polymorphisms (62%) are still subtype-restricted, while only 31% and 7% are type- and supertype-restricted, respectively. These distributions were robust to HLA imputation (see methods and Figure S2). To characterize the extent of differential escape between and within HLA types/subtypes across the HIV-1 proteome, we applied recently developed techniques (see methods and [39]) to explicitly test for differential escape (defined here as cases where one subtype selects a polymorphism and another selects a different — or no — polymorphism at any given codon). Overall, we estimate that when polymorphisms are identified at the type level, there is an 81.5% chance that no differential escape occurs at the subtype level. Conversely, when polymorphisms are identified at the subtype level, there is an 81.2% chance that the subtype of interest indeed displays a significant differential escape pattern compared to the type as a whole (data not shown). Given our 20% false discovery rate significance threshold, results indicate that our method of reporting associations at the level of resolution yielding the lowest p-value is broadly appropriate. Stratifying analyses by individual HLA alleles revealed that the heterogeneous B*15 type exhibited the most evidence for differential escape (with over 85% of pathways being unique at the subtype level) while B*07 exhibited the least (with only 20% of pathways being unique at the subtype level) (Figure 2.2a). Overall, differential CTL escape was widespread at the subtype level, with allele group members exhibiting distinct escape mutations in 58% of the cases on average (HLA-A: 63%, HLA-B: 55%, HLA-C: 60%).

The underlying mechanisms that drive differential CTL escape are unclear, and may include differences in epitope binding, T-cell receptor (TCR)-HLA-epitope interactions, TCR repertoire, or may reflect differences that result from linkage of certain HLA subtypes to external factors, such as antigenic exposure, antigenic processing, or linkage with other HLA class I or class II alleles [39]. In support of a mechanism that

involves properties of the HLA alleles themselves, we observed a striking positive correlation between the overall genetic divergence of HLA exons 2 and 3 (the major genetic determinants of the peptide-HLA binding groove) between allele group members (Figure S3), and the frequency of differential escape between them (Spearman R=0.60, p=0.0043; Figure 2.2b).

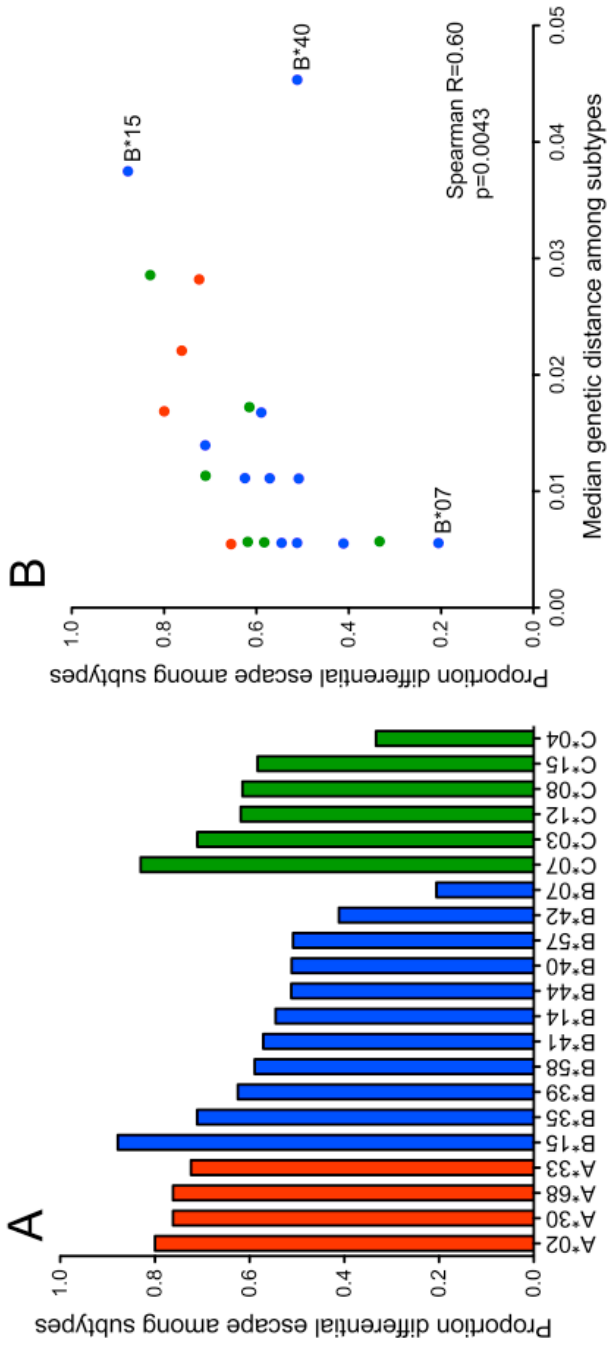

Figure 2.2 Proportion of differential escape amongst HLA subtypes

Widespread differential escape between HLA allele group members correlates with evolutionary distance between HLA genes. Individual HLA-A, -B, and –C types are indicated. (A) The proportion of differential escape (defined as the proportion of HLA-associated polymorphisms for which the odds of escape differed significantly between at least one HLA subtype and the rest of the group) differs markedly for various common types of the HLA-A, -B, and –C loci. (B) HLA types exhibiting the highest level of differential escape were those exhibiting the greatest intratype genetic diversity. doi:10.1128/JVI.01998-12.

2.3.3. Enrichment of escape at epitope-HLA anchor residues

We next examined the distribution and positioning of HLA-associated polymorphisms with respect to CTL epitopes [63]. Our initial epitope reference list consisted of all optimally described CTL epitopes restricted by one or more HLA class I alleles. We then expanded this list by assigning the published epitope to all allele group members (*i.e.* subtypes) belonging to the HLA allele group (*i.e.* supertype, type or serotype, depending on the published restriction) to which the original published epitope belonged (see methods and Table S1). For example, an epitope originally defined as B*57:01-restricted was assigned to all members of the B58 supertype. Since all epitopes on our "expanded" list were defined at subtype level resolution, we utilized the subtype level HLA-associated polymorphisms list in the following analysis, although sensitivity analyses confirmed that results were consistent regardless of the associations list used (not shown).

Of the 2161 HLA-restricted viral escape pathways identified at subtype-level resolution, 29% occurred inside or within +/-3 residues of an optimally described CTL epitope. This value was 21.5% when epitope expansion was not undertaken. In support of including flanking regions in the epitope definition, a bootstrap analysis (100,000 replicates) indicated that escape frequency at epitope flanking residues was significantly higher than escape outside epitopes and their immediate flanking regions for HLA-B alleles ($p=0.00012$), though this was not significant for HLA-A or C alleles. Because the majority of defined epitopes are 9-mers, we first examined these. Of the 262 unique 9-mer HLA-epitope pairs (defined as epitopes restricted by the same HLA supertype and mapping to the same HIV co-ordinates) for which an HLA-associated polymorphism was identified, positions two, three and the C-terminus represented the most frequently escaping sites, while escape at N-terminal flanking residue -3 and C-terminal flanking residue +2 were less frequent than expected under the null hypothesis of equal escape frequency across all sites within/flanking epitopes (binomial test for departure from expected 1/15 frequency, all $p<0.01$, Figure 3A). Stratification by locus revealed that HLA-A-restricted epitopes tended to select for escape mutations at positions 3 and C-terminus, HLA-B restricted epitopes typically developed escape mutations at position 2, and the limited number of HLA-C epitopes most often encoded C-terminal escape mutations (Figure S4).

Classification of escape mutations with respect to their position at an anchor residue (defined according to HLA subtype-specific anchor motifs available at http://www.hiv.lanl.gov/content/immunology/motif_scan/motif.html [76-79]) revealed that escape occurred at anchor residues 1.8-fold more frequently than expected under the null hypothesis of equal escape frequency across all sites within/flanking epitopes (binomial test for departure from expected 2/15 frequency, $p=1.0x10^{-5}$, Figure 2.3b) and 1.4-fold more frequently than expected across all sites within the epitope only (binomial test for departure from expected 2/9 frequency, p=0.003, not shown). The frequency of escape at N- and C-terminal epitope flanking residues was nearly two-fold lower than that expected under the null hypothesis of equal escape frequency across all sites within/flanking epitopes (binomial test for departure from expected 6/15 frequency, $p<1.0x10^{-5}$, Figure 3B). Moreover, the average statistical strength of escape at anchor sites (measured as the maximum absolute natural-log transformed odds ratio [lnOR] for each unique HLA-HIV codon pair) was modestly yet significantly increased compared to that of non-anchor or flanking sites (Median lnOR 2.15 [IQR 1.15-2.58] vs. 1.62 [0.93-2.35] vs. 1.30 [0.81-1.98] respectively, Kruskal-Wallis p=0.035; Figure 2.3c).

Epitopes of other lengths (8, 10 and 11-mer; N=201) were also examined. To ensure alignment of anchor residues, positions were analyzed with respect to their distance from the N and C termini. Relatively uniform frequency of escape was observed at epitope positions 1-4, penultimate and C-terminal positions, while escape at position C-minus 3 (three positions upstream of the C-terminus) was relatively rare (Figure 2.3d). When analyzed separately, 10-mers tended to escape more frequently at position three but not the C-terminus, while 11-mers tended to escape more frequently at position 4 (Figure S4), an observation that may be due to the bulging of oversized epitopes and its effect on TCR contacts [80]. Overall, a 1.45-fold enrichment of escape mutations at anchor residues was observed compared to other sites within and flanking the epitopes (p=0.014; Figure 2.3e); however this did not remain significant when restricted to within-epitope sites only (p=0.3, not shown). The average statistical strength of escape (lnOR) did not differ significantly between anchor and non-anchor sites for 8-, 10- and 11-mer eptopes (Median lnOR was between ~1.4-1.5 for all positions, p=0.89; Figure 2.3f).

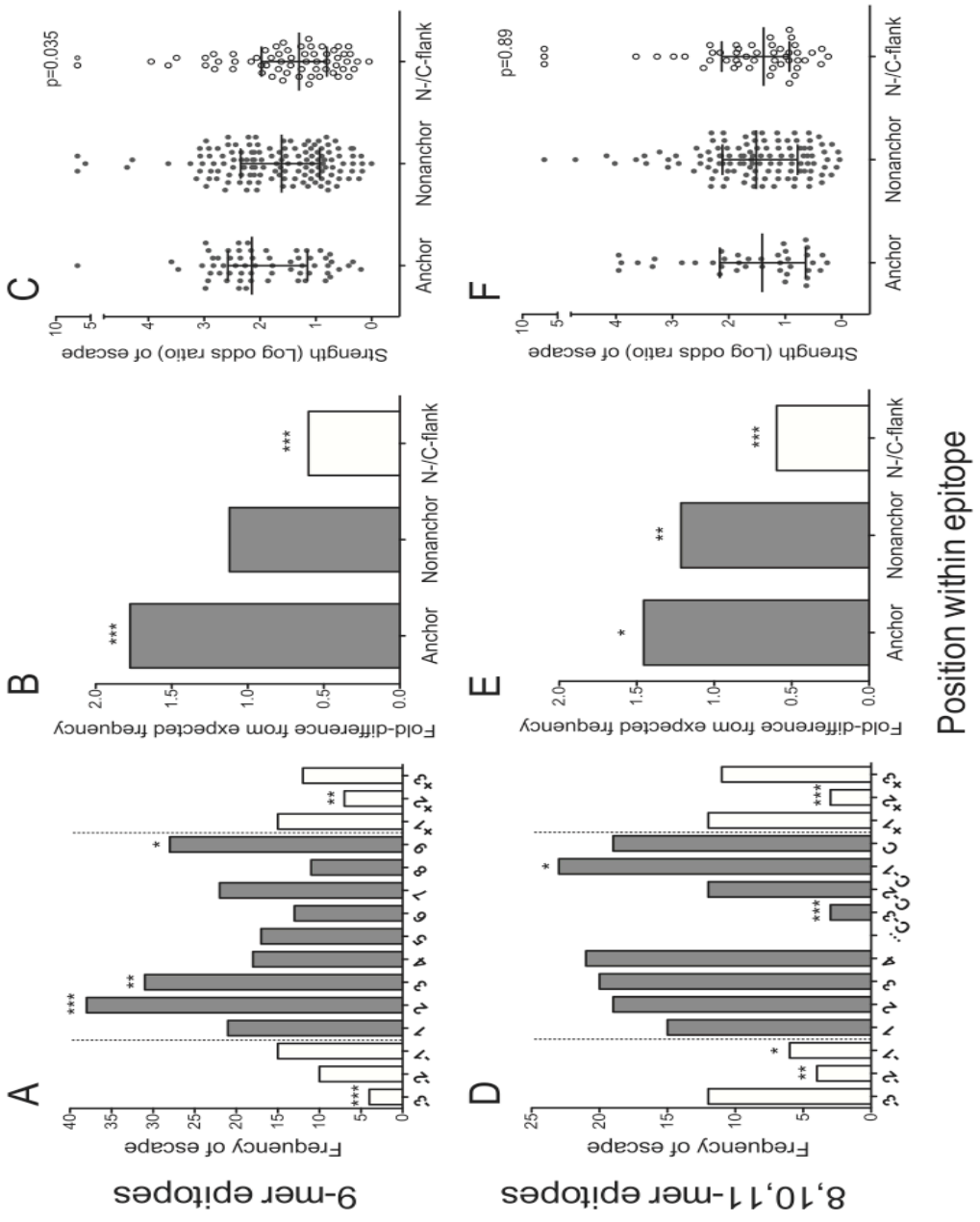

Figure 2.3 The distribution and odds of escape within known epitopes

The distribution and natural log odds ratio of escape within known epitopes is biased towards anchor residues. The distribution and natural log odds ratio of escape mutations within known CTL 9-mer epitopes (A, B, and C) and 8-/10-/11-mer epitopes (D, E, and F) are shown. Frequencies of escape by relative position within or occurring at the +- 3 amino acids flanking the epitope termini are depicted as histograms in panels A and D. In panel D, notation of the type C-3 on the x-axis indicates the number of positions upstream of the C-terminus (in this case, 3). These data are depicted as fold differences from expected frequency of escape occurring at HLA-specific anchor, nonanchor, and epitope flanking positions (null hypothesis of equal escape probability across all positions) and are shown in panels B and E. *, $P < 0.05$; **, $P < 0.001$; ***, $P < 0.0001$ (binomial test). The natural log of the odds ratio of escape at HLA-specific anchor, nonanchor, and epitope flanking positions are shown as scatter plots (C and F). doi:10.1128/JVI.01998-12.

2.3.4. *Estimating the effect of anchor residue escape on HLA binding*

Given the enrichment of CTL escape mutations at HLA anchor residues, we next estimated the extent to which these mutations were predicted to affect peptide-HLA binding using NetMHCpan2.4, an artificial neural network-based peptide-HLA binding prediction tool trained on >37,000 quantitative binding data for more than 42 different HLA molecules [81]. For each HLA-associated "adapted" polymorphism that occurred inside an epitope (N=147 9-mers and N=133 8/10/11-mers), we compared the predicted HLA binding affinity for the published epitope in its HLA-restricted "nonadapted" (*i.e.* susceptible) versus its HLA-associated "adapted" (*i.e.* escaped) form, using the subtype B consensus for the epitope backbone (in cases where no HLA-restricted "nonadapted" residue was identified, the subtype B consensus residue was used). After removing epitopes predicted to bind the restricting HLA with IC_{50}>1000 nM in their "nonadapted" form (representing a 2-fold lower affinity than NetMHCpan2.4-defined weak binding threshold; which included N=30 9-mers and N=38 8/10/11-mers), we calculated that anchor residue escape in 9-mer epitopes was predicted to reduce peptide-HLA binding a median of 8.9-fold (IQR 3.7-52.1-fold) compared to 1.25-fold (IQR 0.80-1.99-fold) at non-anchor sites (Figure 2.4a). Among 8/10/11-mer epitopes, anchor residue escape was predicted to reduce peptide binding a median of 8.9-fold (IQR 1.9-114.3) compared to 1.0 -fold (IQR 0.87-1.56) at non-anchor sites (Figure 2.4b). Overall, 36 out of 212 (17%) within-epitope escape mutations were predicted to "completely" abrogate peptide-HLA binding, defined as epitopes which bound with an IC_{50}<1000 nM affinity in their nonadapted form but IC_{50}>1000 nM affinity in their escaped form (not shown).

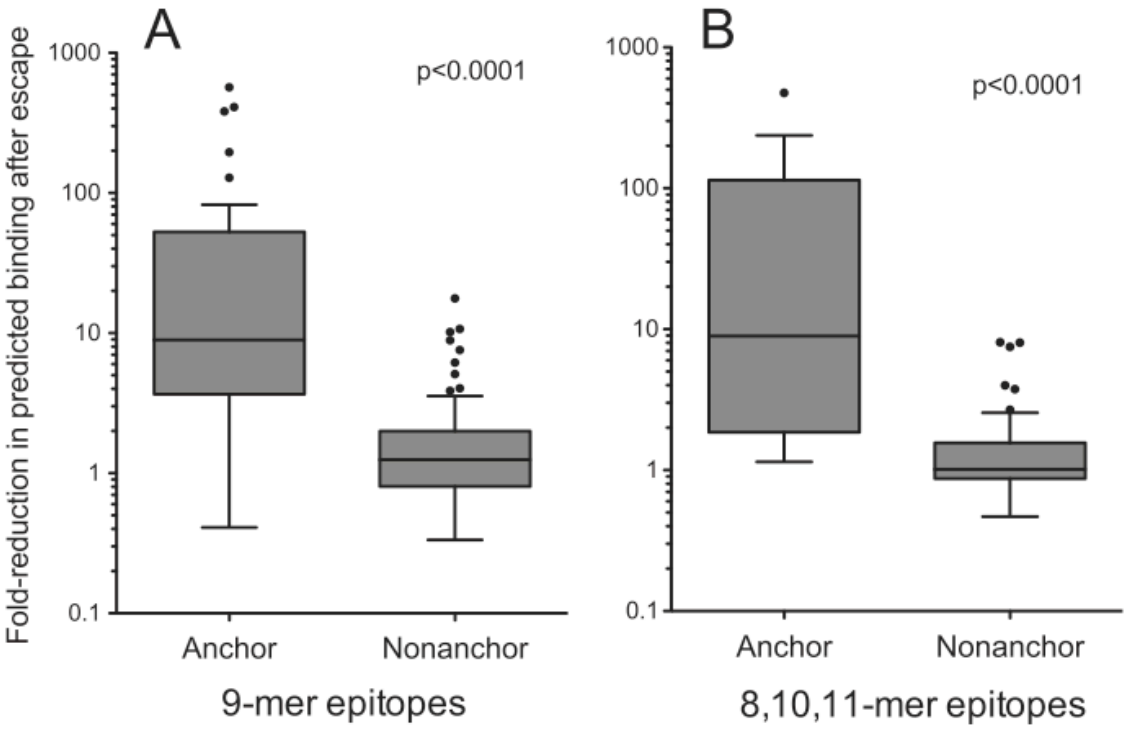

Figure 2.4 Change in predicted peptide-HLA binding affinities

Fold changes in predicted peptide-HLA binding affinities following escape at anchor residues. A) Fold changes in predicted peptide-HLA binding affinities as a consequence of escape at HLA-specific anchor (versus nonanchor) residues within 9-mer. B) The same analysis within 8-, 10-, and 11-mer epitopes. doi:10.1128/JVI.01998-12.

2.3.5. Correlates of protective immunity revealed by analysis of population level immune escape pathways

Expression of specific HLA class I alleles is associated with differential rates of HIV-1 disease progression [73,82,83], but the precise mechanisms underlying these effects remain incompletely known [84-87]. Given that HLA-associated polymorphisms identify viral sites under strong *in vivo* selection by an HLA allele at the population level, analysis of their frequency, distribution, statistical strength, sequence conservation and other characteristics allows us identify characteristics that discriminate protective HLA allele-associated immune pressures from non-protective ones. It is important to emphasize that, although population-level HLA-associated polymorphisms may allow us to infer specific properties, tendencies or attributes that render certain HLA alleles particularly effective at controlling HIV, by no means are we proposing that selection of escape mutations is beneficial at the individual patient level.

HLA-associated protective effects were defined according to published hazard ratios for progression to AIDS (HR-AIDS) derived from a large longitudinal natural history cohort (see methods and [73]). We investigated 33 interrelated features of viral sites under HLA-mediated selection as potential correlates of protective immunity. These 33 features were divided into two broad categories: those related to their frequency/distribution (overall, and within known epitopes), and those related to the statistical strength of selection and/or mutational constraints on the sites themselves (Table 2.1). Where appropriate, analyses were undertaken at proteomewide and individual-protein levels. HLA-associated polymorphisms from the subtype-level analysis were used.

Table 2.1 Identifying correlates of HLA associated protective immunity

Feature	Category	No. of HLA alleles analyzed	Spearman *R*	*P* value	*q* value
Median OR of selection in Gag	Strength/constraint	26	−0.60	0.0012	**0.04**
No. of sites under selection (proteome-wide)	Frequency/distribution	53	−0.41	0.0024	**0.04**
Proportion of selected sites occurring at HLA anchor residues (proteome-wide [%])	Frequency/distribution	34	−0.47	0.0052	**0.04**
No. of sites under selection in Gag	Frequency/distribution	53	−0.38	0.0054	**0.04**
No. of sites under selection in accessory proteins	Frequency/distribution	53	−0.37	0.0058	**0.04**
No. of selected sites occurring at HLA anchor residues (proteome-wide)	Frequency/distribution	44	−0.35	0.018	**0.10**
No. of sites under selection within/near epitopes (proteome-wide)	Frequency/distribution	53	−0.31	0.025	**0.11**
No. of active epitopes in Gag	Frequency/distribution	53	−0.30	0.027	**0.11**
Median no. of sites under selection per epitope in Gag	Frequency/distribution	22	−0.46	0.032	**0.12**
Median OR of selection in Pol	Strength/constraint	38	−0.34	0.039	**0.13**
No. of sites under selection in Env	Frequency/distribution	53	−0.28	0.045	**0.13**
No. of sites under selection in Pol	Frequency/distribution	53	−0.28	0.046	**0.13**
Median conservation of selected sites in Nef	Strength/constraint	38	−0.29	0.079	0.20
Median conservation of selected sites in accessory proteins	Strength/constraint	45	0.24	0.11	0.25
Median conservation of selected sites in Env	Strength/constraint	32	0.28	0.13	0.28
No. of ites under selection within/near epitopes in Gag	Frequency/distribution	28	−0.27	0.16	0.33
No. of active epitopes (proteome-wide)	Frequency/distribution	53	−0.17	0.23	0.41
Median no. of covarying codons per selected site (proteome-wide)	Strength/constraint	51	−0.17	0.23	0.41
Median no. of sites under selection per epitope (proteome-wide)	Frequency/distribution	44	−0.18	0.24	0.41
Median conservation of selected sites in Nef	Strength/constraint	51	0.16	0.25	0.42
Proportion of sites under selection in Gag (%)	Frequency/distribution	51	−0.14	0.33	0.50
Proportion of sites under selection in Env (%)	Frequency/distribution	51	−0.13	0.35	0.50
Proportion of sites under selection in Nef (%)	Frequency/distribution	51	0.13	0.36	0.50
Proportion of sites under selection in accessory proteins (%)	Frequency/distribution	51	−0.13	0.37	0.50
No. of sites under selection in Nef	Frequency/distribution	53	−0.12	0.39	0.50
Median OR of escape in Nef	Strength/constraint	38	0.14	0.40	0.50
Median no. of covarying codons per site under selection (Gag only)	Strength/constraint	26	−0.084	0.68	0.81
Median OR of selection in accessory proteins	Strength/constraint	45	0.059	0.70	0.81
Median OR of selection in Env	Strength/constraint	32	−0.067	0.72	0.81
Proportion of sites under selection in Pol (%)	Frequency/distribution	51	−0.0069	0.96	0.99
Median conservation of selected sites in Gag	Strength/constraint	26	−0.0068	0.97	0.99
Median OR of selection (proteome-wide)	Strength/constraint	50	−0.0024	0.99	0.99
Median conservation of selected sites in Pol	Strength/constraint	39	0.0016	0.99	0.99

Of the 33 variables investigated, we identified 12 potential correlates of HLA-mediated protective immunity against disease progression ($p<0.05$, $q<0.2$) (Table 1). Among the strongest correlates of protection was the significant inverse relationship between total number of sites under selection by a given HLA and its HR-AIDS (Spearman's R=-0.41; p=0.0024; Figure 2.5a), indicating that protective HLA alleles exert substantial *in vivo* pressure on a greater number of sites across the HIV proteome. This observation was almost exclusively driven by HLA-A and B alleles (Spearman's R=-0.48; p=0.0016; not shown). When analyzed at the individual protein level, this inverse relationship was strongest for Gag and remained significant for all proteins except Nef, where the trend persisted but was not significant (Table 2.1).

Notably, no significant correlations were observed between HR-AIDS and median sequence conservation of sites under selection (where the sequence conservation of a given site was defined as the proportion of the total cohort harbouring the consensus residue at this position [50]). This remained true when the bias induced by HLA-mediated selection on these sites was addressed by excluding persons expressing these alleles from the calculation of conservation (Table 2.1). Similarly no significant correlations were observed between HR-AIDS and the median number of covarying codons per site under selection. However, significant inverse relationships were observed between an HLA allele's HR-AIDS and the median odds ratio of selection for mutations in Gag (R= -0.6, p=0.0012; Figure 2.5b), and to a lesser extent in Pol (R= -0.34, p=0.039), but not overall or in other HIV-1 proteins.

Another highly significant proteomewide correlate of protection was the tendency of protective alleles to select escape mutations at HLA anchor sites within known epitopes (Spearman's R=-0.47; p=0.0052; Figure 2.5c). This effect appeared to be largely constrained to Pol and Gag (p=0.0015 and p=0.065, respectively). The median number of sites under selection per epitope in Gag, but not elsewhere in the viral genome, was also modestly inversely correlated with HR-AIDS (Table 2.1).

To attempt to identify which of these features represented the strongest independent correlates of protection, we investigated a variety of multivariate models incorporating four of our top "hits" (Median OR of selection in Gag, Median OR of selection in Pol, # sites under selection proteomewide and % selected sites occurring at anchor residues). Unfortunately due to the strong interdependency of these variables

and the fact that not all variables were available for all HLA alleles, multivariate analyses were not robust to changes in model selection procedures. The only variable that consistently emerged as a significant independent predictor of protection was the number of sites under selection by a given HLA across the HIV-1 proteome (not shown).

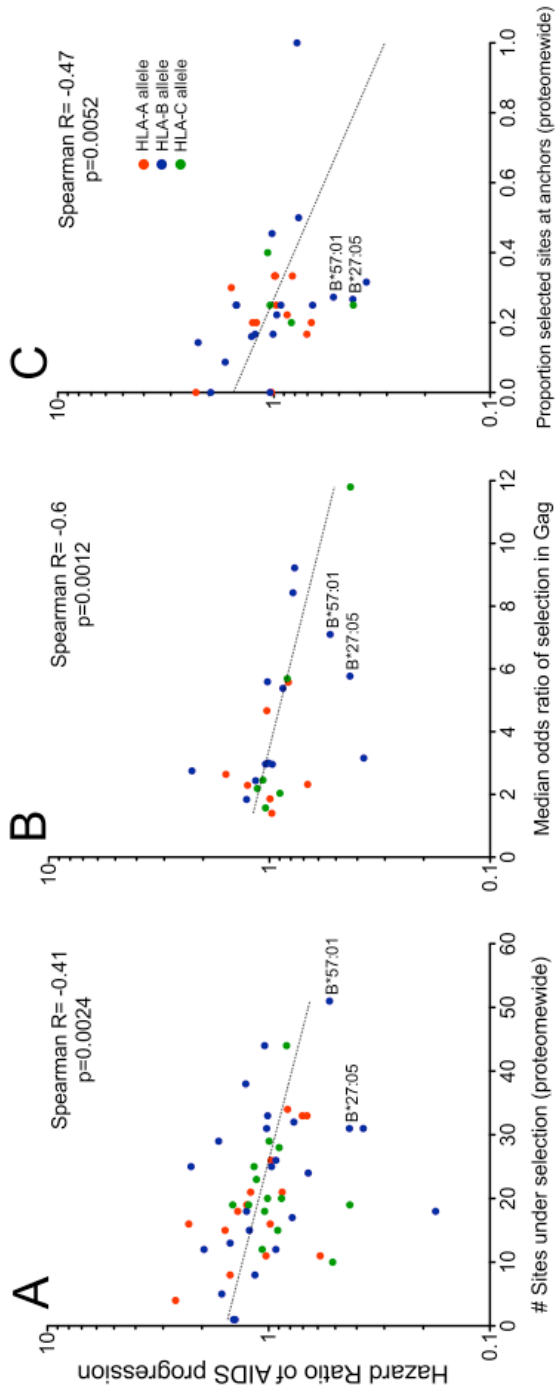

Figure 2.5 Immune escape pathways reveal correlates of protective immunity

Population-level immune escape pathways in HIV-1 reveal correlates of protective immunity: selected results. Individual HLA-A, -B, and –C alleles are as indicated. Significant inverse correlations were observed between A) the total numbers of sites under immune selection by a given HLA across the entire HIV-1 proteome, B) the median odds ratio of escape in Gag, C) and the percent escape at HLA anchor residues and published hazard ratios of AIDS progression.

doi:10.1128/JVI.01998-12.

2.4. Discussion

HLA-associated polymorphisms identify HIV-1 sites under consistently strong immune pressure *in vivo* and whose sequence diversity is largely driven by HLA. As such, the well-powered nature of this analysis combined with its proteome-wide focus allowed us to study the frequency, distribution and characteristics of viral sites under HLA-mediated selection. The proportion of population-level sequence variation attributable, at least in part, to HLA (defined as the proportion of sites in a given protein harbouring at least one HLA-associated polymorphism) differs markedly by protein, and is independent of the protein's overall conservation. For example, whereas the majority of variable Nef codons are attributable, at least in part, to HLA class I-mediated pressures, the latter explain only a quarter of Vpu sequence variation, implying other host factors in driving evolution of this viral gene[88-90] (to provide context, nearly 40% of variable sites in the highly conserved $p24^{Gag}$ protein are associated with HLA).

Building on recent advances[38,39], we demonstrated significant variation in the proportion of differential escape among closely related HLA subtypes, implicating differences in TCR repertoire and/or the interaction between TCR and the epitope-HLA complex as the underlying mechanisms in at least some cases[91]. Widespread differential escape is consistent with reported differences in HIV-1 disease progression between HLA alleles differing by as little as one amino acid[37,92,93], and also supports the notion that analysis of HLA-associated polymorphisms provides a novel perspective from which to discriminate both quantitative and qualitative differences in HLA-restricted immune responses. Of interest, HLA allele subtype pairs that exhibit the highest level of differential escape were those with the greatest intratype sequence divergence, raising the intriguing hypothesis that differential escape may be a consequence of active positive selection at these loci.

Analysis of the distribution of HIV-1 escape mutations within known CTL epitopes revealed that abrogation of peptide-HLA binding via anchor residue mutation represented a major escape mechanism, and that such mutations incurred, on average, a nearly tenfold reduction in predicted HLA binding affinity. Among 9-mer epitopes, the statistical strength of escape mutations at anchor residues was higher than for non-anchor residues. Mutations occurring at non-anchor positions within the epitope were less frequently detected, and those at residues immediately flanking the N- and C-

termini of the epitope were less frequent still (although still significantly higher than in extra-epitopic zones, at least for HLA-B). Although it is tempting to conclude that abrogation of peptide-HLA binding represents HIV's favoured escape mechanism, a potential limitation should be noted. It is possible that detection of HLA-associated polymorphisms at the population level is inherently biased towards detection of "straightforward" escape mutations such as those directly affecting binding to HLA, while escape mutations resulting from more complex and/or "secondary" interactions (*e.g.:* escape from populations of TCR capable of interacting with HLA-bound peptide) may be more challenging to detect. Indeed, numerous examples exist where an escape variant selected in one individual is highly recognized by another's immune response (e.g.: [94-97]).

Despite these caveats, recent large-scale studies of immune escape indicate that datasets of the present size have sufficient power to detect even rare escape mutations [39], as well as to detect cases where escape from one HLA allele represents the susceptible form for another[18]. The high frequency of anchor residue escape has implications for proposed vaccine strategies aimed at immunizing with both wild-type and common variants to channel HIV evolution down unconventional pathways [16,18,48]. While still valid, such strategies should consider excluding epitope variants that are experimentally verified to not bind their restricting HLA, unless such variants overlap key epitopes restricted by other alleles and/or lead to the creation of novel epitopes relevant to immune control[52].

Analysis of HLA-associated polymorphisms identified at the population level also provide a unique perspective from which to identify, albeit indirectly, specific characteristics that differentiate protective from non-protective HLA-restricted CTL responses. Such sequence-based methods offer an alternative perspective than traditional *in vitro* and/or *ex vivo* assays such as IFNg ELISPOT[98], where, for a variety of reasons (including use of synthetic consensus peptides at supra-physiological concentrations that bypass antigen processing machinery, detection of a single cytokine readout only, inability to detect previous effective responses due to escape in the autologous virus, and others[99]), responses detected *in vitro* may not accurately discriminate effective responses occurring *in vivo.* Indeed, a recent large-scale analysis of >2000 chronically clade C-infected persons identified differential immune escape as a better predictor of average per-HLA plasma viral load than responses measured by IFNg ELISPOT[39].

Importantly, analysis of population-level HLA-associated polymorphisms as markers of *in vivo* immune pressure by no means argues that CTL escape is protective at the individual level — indeed, escape mutations in individuals have been linked to higher viral load and disease progression[39,41,100,101]. At the population level, HLA-associated polymorphic sites identify HLA-restricted CTL responses that are consistently mounted in a substantial number of individuals expressing that allele—responses that are strong enough to drive the virus to escape in reproducible ways. As such, these sites represent the total potential of common CTL responses restricted by a given HLA allele to effectively target HIV; analysis of their distribution and characteristics can therefore help illuminate correlates of protection. Translation of these findings to the individual level is more complex. First, not all individuals expressing a given HLA will mount all possible CTL responses. Broadly generalizing, from a viral control perspective the ideal would be to mount an effective immune response that the virus is incapable of escaping. Less ideal (but more the norm) would be to generate an effective response that the virus ultimately is able to escape (yielding losses of immune control that in some cases may be offset by viral fitness costs). Least ideal would be the inability to mount the response in the first place. At the individual level therefore, escape mutations can be regarded as genetic relics of formerly effective CTL responses, where, in general, the negative consequences of escape are still preferable to the inability to mount responses at all[39].

Our proteome-wide and protein-specific analyses shed new light on the longstanding debate around whether CTL response breadth is[102,103] or is not[104] a correlate of protection. For example, the number of Env-specific CD8 T-cell responses in chronic infection, as measured by IFNg ELISPOT, has been associated with poorer HIV control at the individual level[105,106], suggesting that not all CTL responses are equally effective[31,107,108]. Our results indicate that protective alleles (most notably those of the HLA-A, and B loci) have the potential to exert immune pressure on a larger number of sites across the viral proteome, and that selection breadth potential is significantly associated with protection for all proteins except Nef (although a trend remained for the latter). Furthermore, the number of associated sites across the proteome per HLA allele was the only variable that consistently emerged as an independent correlate of protection in multivariate analyses. Our observations highlight the utility of using population-level HLA-associated polymorphisms to preferentially discriminate the impact

of HLA-restricted immune responses that are most effective at controlling HIV-1 replication *in vivo*, and suggest that the ability of an HLA allele to mount broad CTL selection pressures to a wide variety of proteins beyond Gag — including Env — represents a general correlate of protection.

The negative results of this study are also illuminating. The lack of correlation between HR-AIDS and density of the amino acid covariation network associated with any given site under immune selection suggests that targeting structural or functional "network hubs" (*i.e.* those exhibiting higher-order evolutionary constraints within a given HIV protein [109]) is not independently protective. Similarly, the lack of correlation between HR-AIDS and average sequence conservation of sites under HLA selection indicates that the ability to target constrained sites is not in itself protective: it may depend on where these sites are in the HIV-1 proteome and how strongly they are targeted. Indeed, analysis at the individual protein level revealed that the statistical strength of selection on conserved sites (as measured by odds ratio of association) in Gag, and to a lesser extent Pol, but not other viral proteins, was a significant correlate of protection. Since a high odds ratio reflects both the frequency of selection among persons expressing the HLA allele and the likelihood of reversion in persons who lack the allele, it may be interpreted as an index of exceptionally strong selection at mutationally constrained sites. Our results thus highlight strength of immune selection on conserved sites *within structural and functional HIV-1 proteins* as an important correlate of protection.

Gag was also confirmed as a particularly effective immune target. The number of sites under active immune selection in Gag, the number of "active" Gag epitopes (*i.e.* those harbouring evidence of immune selection) and the average number of sites under selection per Gag epitope (a marker of diversity of selection pressures on a given epitope) also represented significant correlates of protection. Although an HLA allele's ability to consistently mount immune pressure on conserved Gag codons may appear an intuitive correlate of protection, the factors that constrain the virus to escape at those specific sites remain unclear. That the median odds ratio of escape in Gag correlated significantly with both the number of sites under selection and the proportion of anchor residue escape in the same protein, and that in 9-mers the odds ratio of escape was significantly higher at anchor sites, suggests that a hallmark of HLA-associated protection is the ability to consistently mount intense CTL responses from which the

virus is compelled to escape (hence the high odds of escape), but can only do so at a limited number of positions (hence the position specificity) and at a biologically relevant fitness cost (hence sequence conservation in persons lacking the allele). Furthermore, if the anchor residue is otherwise highly conserved, escape at this position could indicate that CTL targeting the epitope are highly polyclonal and/or cross-reactive, thus limiting escape options to those that abrogate epitope presentation. If so, this supports the clonal composition and/or diversity of epitope-specific CTL repertoire as a correlate of protection[31,110], and supports vaccine strategies seeking to stimulate CTL responses to both wild-type and commonly occurring escape variants at non-anchor sites[111-113], most notably in epitopes where escape at anchor residues is likely to incur a fitness cost. Indeed, a recent analysis limited to Pol epitopes suggested that mutations which disrupt peptide binding to protective HLA alleles (particularly HLA-A alleles) may be more fitness-costly than escape from non-protective HLA alleles[114].

Although results provide potentially useful information to guide future vaccine and immunogenicity studies, some limitations merit mention. We employed bioinformatics and statistical approaches to infer general mechanisms of immune evasion and its consequences for epitope presentation *in vivo*, but individual predictions will require rigorous experimental verification prior to consideration as potential immunogens. Furthermore, although well-powered analyses of HLA-associated polymorphisms identify the majority of escape pathways commonly driven by a given HLA[39], it may still underestimate the amount of immune selection imposed on HIV at the population level. For example, as our method only detects HIV polymorphisms whose frequencies differ statistically among persons having or lacking a particular HLA allele, it is unable to detect immune pressures directed against viral codons so highly conserved that the benefits of escape would not outweigh the fitness costs. The B*57-associated KF11 epitope provides a potential example of such a phenomenon: when presented by B*57:01 (but not B*57:03) in context of subtype B, a highly conserved, cross-reactive TCR repertoire is induced, from which the virus escapes with great difficulty[91]. Furthermore, although such well-powered analyses can identify extremely rare pathways [39], they may still underestimate the amount of immune selection imposed on HIV at the individual level; indeed, the extent to which unique immune responses (and their escape pathways) occur, and affect disease progression, remain unclear. Our method may also preferentially detect “straightforward” escape pathways (such as those directly impacting

peptide binding to HLA) over those that may be more "multifactorial" and/or variable among individuals (e.g.: TCR escape that could occur at a variety of intra-epitope positions depending on the individual's T-cell repertoire); the extent of this potential bias could be quantified by longitudinally analyzing escape sites in seroconverter cohorts (e.g.:[40,94,115]). Moving forward, it will be essential to translate indirect correlates of protection at the HLA allele level to specific features of protective HLA-restricted CTL responses at the individual level.

The present study confirms the differential influence of HLA class I-mediated selection pressures on population-level sequence diversity of individual HIV-1 proteins, and the common occurrence of differential escape among closely related alleles[17-19,21,39,57]. That the majority of HLA-associated polymorphisms occur outside known epitopes suggests that many epitopes remain undiscovered; indeed HLA-associated polymorphisms represent an excellent tool to guide epitope discovery as they are unbiased by consensus sequences or limited knowledge of binding motifs (*e.g.*: [16,42,116,117]). This study also extends our understanding of the proportion of population-level HIV-1 diversity attributable to HLA selection pressures, identifies abrogation of HLA-peptide-binding as a predominant escape mechanism, suggests a potential evolutionary mechanism underlying differential escape between allele group members, and provides an extensive reference of proteome-wide HLA-mediated escape pathways.

Importantly, our hypothesis that HLA-associated polymorphisms mark sites under active immune selection by HLA alleles *in vivo* affords a novel perspective from which to identify features of such responses potentially relevant to vaccine design. Specifically, the tendency for protective alleles to exert selection pressures on a larger number of viral sites highlights selection breadth as a general correlate of protection. The tendency of HIV-1 to evade immune pressure by protective HLA alleles via mutations at anchor residues and/or at multiple sites within epitopes supports CTL repertoire composition and/or diversity as a correlate of protection[31,110], although this remains controversial[118]. The average sequence conservation of sites under selection was not a general correlate of protection, indicating that the factors underpinning protective immune responses are more complex than simply the ability to target conserved regions. However, that the ability to *strongly* mount immune pressure on conserved sites in Gag and Pol was protective extends the evidence supporting Gag as a particularly effective immune target [105,119-121] and suggests that Gag (and Pol)-derived peptides possess characteristics

which render them inherently more effective as immune targets: rapid presentation of Gag-, and to a lesser extent Pol-, derived peptides from incoming virions[122,123] provides an intriguing explanation for this phenomenon. Taken together, our results suggest that vaccine strategies aimed at stimulating broad CTL responses towards non-Nef epitopes, and in particular, intense polyclonal responses that would force escape at conserved anchor positions, especially within constrained Gag and Pol epitopes, may be particularly effective at controlling HIV-1. Analyses of population-level escape pathways for other viral pathogens (such as hepatitis C virus)[124-126] may similarly illuminate relevant correlates of protective cellular immunity.

2.5. References

1. Carlson JM, Brumme CJ, Martin E, et al. Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. *Journal of virology.* Dec 2012;86(24):13202-13216.

2. Nickle DC, Rolland M, Jensen MA, et al. Coping with viral diversity in HIV vaccine design. *PLoS Comput Biol.* Apr 27 2007;3(4):e75.

3. Gaschen B, Taylor J, Yusim K, et al. Diversity considerations in HIV-1 vaccine selection. *Science.* Jun 28 2002;296(5577):2354-2360.

4. Taylor BS, Hammer SM. The challenge of HIV-1 subtype diversity. *N Engl J Med.* Oct 30 2008;359(18):1965-1966.

5. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science.* Dec 1 1989;246(4934):1155-1158.

6. Johnson VA, Calvez V, Gunthard HF, et al. 2011 update of the drug resistance mutations in HIV-1. *Top Antivir Med.* Nov 2011;19(4):156-164.

7. Borrow P, Lewicki H, Wei X, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med.* Feb 1997;3(2):205-211.

8. Koup RA. Virus escape from CTL recognition. *J Exp Med.* Sep 1 1994;180(3):779-782.

9. Price DA, Goulder PJ, Klenerman P, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A.* Mar 4 1997;94(5):1890-1895.

10. Phillips RE, Rowland-Jones S, Nixon DF, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature.* Dec 12 1991;354(6353):453-459.

11. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol.* Aug 2004;4(8):630-640.

12. Gnanakaran S, Bhattacharya T, Daniels M, et al. Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS Pathog.* Sep 2011;7(9):e1002209.

13. Rolland M, Tovanabutra S, Decamp AC, et al. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nat Med.* Mar 2011;17(3):366-371.

14. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science.* May 24 2002;296(5572):1439-1443.

15. Allen TM, Altfeld M, Geer SC, et al. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J Virol.* Nov 2005;79(21):13239-13249.

16. Bhattacharya T, Daniels M, Heckerman D, et al. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science.* Mar 16 2007;315(5818):1583-1586.

17. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology.* Nov 2008;4(11):e1000225.

18. Brumme ZL, Brumme CJ, Heckerman D, et al. Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. *PLoS Pathog.* Jul 6 2007;3(7):e94.

19. Rousseau CM, Daniels MG, Carlson JM, et al. HLA Class-I Driven Evolution of Human Immunodeficiency Virus Type 1 Subtype C Proteome: Immune Escape and Viral Load. *J Virol.* Apr 23 2008;82(13):6434-6446.

20. Matthews PC, Prendergast A, Leslie A, et al. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *J Virol.* Jul 2 2008;82(17):8548-8559.

21. Rolland M, Carlson JM, Manocheewa S, et al. Amino-acid co-variation in HIV-1 Gag subtype C: HLA-mediated selection pressure and compensatory dynamics. *PLoS ONE.* 2010;5(9).

22. Brumme ZL, John M, Carlson JM, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PloS one.* 2009;4(8):e6687.

23. John M, Heckerman D, James I, et al. Adaptive interactions between HLA and HIV-1: Highly divergent selection imposed by HLA class I molecules with common supertype motifs. *Journal of Immunology.* 2010;184(8):4368-4377.

24. Matthews PC, Leslie AJ, Katzourakis A, et al. HLA Footprints on HIV-1 Are Associated with Inter-Clade Polymorphisms and Intra-Clade Phylogenetic Clustering. *J Virol.* Feb 25 2009;83(9):4605-4615.

25. Kelleher AD, Long C, Holmes EC, et al. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J Exp Med.* Feb 5 2001;193(3):375-386.

26. Ammaranond P, Zaunders J, Satchell C, van Bockel D, Cooper DA, Kelleher AD. A new variant cytotoxic T lymphocyte escape mutation in HLA-B27-positive individuals infected with HIV type 1. *AIDS research and human retroviruses.* May 2005;21(5):395-397.

27. Draenert R, Le Gall S, Pfafferott KJ, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *The Journal of experimental medicine.* Apr 5 2004;199(7):905-915.

28. Yokomaku Y, Miura H, Tomiyama H, et al. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. *J Virol.* Feb 2004;78(3):1324-1332.

29. Price DA, West SM, Betts MR, et al. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. *Immunity.* Dec 2004;21(6):793-803.

30. Theodossis A, Guillonneau C, Welland A, et al. Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. *Proc Natl Acad Sci U S A.* Mar 23 2010;107(12):5534-5539.

31. Iglesias MC, Almeida JR, Fastenackels S, et al. Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood.* Aug 25 2011;118(8):2138-2149.

32. Cale EM, Bazick HS, Rianprakaisang TA, Alam SM, Letvin NL. Mutations in a dominant Nef epitope of simian immunodeficiency virus diminish TCR:epitope peptide affinity but not epitope peptide:MHC class I binding. *J Immunol.* Sep 15 2011;187(6):3300-3313.

33. Iversen AK, Stewart-Jones G, Learn GH, et al. Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nat Immunol.* Feb 2006;7(2):179-189.

34. Stewart-Jones GB, di Gleria K, Kollnberger S, McMichael AJ, Jones EY, Bowness P. Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. *Eur J Immunol.* Feb 2005;35(2):341-351.

35. Brackenridge S, Evans EJ, Toebes M, et al. An early HIV mutation within an HLA-B*57-restricted T cell epitope abrogates binding to the killer inhibitory receptor 3DL1. *J Virol.* Jun 2011;85(11):5415-5422.

36. Avila-Rios S, Ormsby CE, Carlson JM, et al. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Retrovirology.* 2009;6:72.

37. Ngumbela KC, Day CL, Mncube Z, et al. Targeting of a CD8 T cell env epitope presented by HLA-B*5802 is associated with markers of HIV disease progression and lack of selection pressure. *AIDS Res Hum Retroviruses.* Jan 2008;24(1):72-82.

38. Kloverpris HN, Stryhn A, Harndahl M, et al. HLA-B*57 Micropolymorphism shapes HLA allele-specific epitope immunogenicity, selection pressure, and HIV immune control. *J Virol.* Jan 2012;86(2):919-929.

39. Carlson JM, Listgarten J, Pfeifer N, et al. Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1. *J Virol.* May 2012;86(9):5230-5243.

40. Brumme ZL, Brumme CJ, Carlson J, et al. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol.* Sep 2008;82(18):9216-9227.

41. Brumme ZL, Tao I, Szeto S, et al. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS.* Jul 11 2008;22(11):1277-1286.

42. Almeida CA, Roberts SG, Laird R, et al. Exploiting knowledge of immune selection in HIV-1 to detect HIV-specific CD8 T-cell responses. *Vaccine.* Aug 23 2010;28(37):6052-6057.

43. Berger CT, Carlson JM, Brumme CJ, et al. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *J Exp Med.* Jan 18 2010;207(1):61-75, S61-12.

44. Bansal A, Carlson J, Yan J, et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J Exp Med.* Jan 18 2010;207(1):51-59, S51-53.

45. Leslie A, Kavanagh D, Honeyborne I, et al. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J Exp Med.* Mar 21 2005;201(6):891-902.

46. Kawashima Y, Pfafferott K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature.* Feb 25 2009;458(7238):641-645.

47. Rolland M, Frahm N, Nickle DC, et al. Increased Breadth and Depth of Cytotoxic T Lymphocytes Responses against HIV-1-B Nef by Inclusion of Epitope Variant Sequences. *PLoS ONE.* 2011;6(3):e17969.

48. Altfeld M, Allen TM. Hitting HIV where it hurts: an alternative approach to HIV vaccine design. *Trends Immunol.* Nov 2006;27(11):504-510.

49. Chopera DR, Wright JK, Brockman MA, Brumme ZL. Immune-mediated attenuation of HIV-1. *Future Virol.* Aug 2011;6(8):917-928.

50. Rolland M, Nickle DC, Mullins JI. HIV-1 group M conserved elements vaccine. *PLoS Pathog.* Nov 2007;3(11):e157.

51. Mallal SA. The Western Australian HIV Cohort Study, Perth, Australia. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1998;17 Suppl 1:S23-27.

52. Almeida CA, Bronke C, Roberts SG, et al. Translation of HLA-HIV associations to the cellular level: HIV adapts to inflate CD8 T cell responses against Nef and HLA-adapted variant epitopes. *J Immunol.* Sep 1 2011;187(5):2502-2513.

53. Haas DW, Wilkinson GR, Kuritzkes DR, et al. A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. *HIV Clin Trials.* Sep-Oct 2003;4(5):287-300.

54. Woods CK, Brumme CJ, Liu TF, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J Clin Microbiol.* Mar 7 2012.

55. Cotton LA, Rahman MA, Ng C, et al. HLA class I sequence-based typing using DNA recovered from frozen plasma. *J Immunol Methods.* Aug 31 2012;382(1-2):40-47.

56. Listgarten J, Brumme Z, Kadie C, et al. Statistical resolution of ambiguous HLA typing data. *PLoS Comput Biol.* Feb 2008;4(2):e1000016.

57. Carlson J, Kadie C, Mallal S, Heckerman D. Leveraging hierarchical population structure in discrete association studies. *PLoS ONE.* 2007;2(7):e591.

58. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological).* 1977;39(1):1-38.

59. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America.* Aug 5 2003;100(16):9440-9445.

60. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. *Nucleic Acids Res.* Jan 2011;39(Database issue):D1171-1176.

61. *PHYLIP (Phylogeny Inference Package) version 3.6.* [computer program]. Seattle: Department of Genome Sciences, University of Washington.; 2005.

62. Veerassamy S, Smith A, Tillier ER. A transition probability model for amino acid substitutions from blocks. *J Comput Biol.* 2003;10(6):997-1010.

63. Llano A, Frahm N, Brander C. How to optimally define optimal cytotoxic T lymphocyte epitopes in HIV infection? In: Yusim K, Korber BT, Brander C, et al., eds. *HIV Molecular Immunology*. Los Alamos, New Mexico: Los Alamos National Laboratory, Theoretical Biology and Biophysics; 2009:3-24.

64. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC immunology.* 2008;9:1.

65. Goulder PJ, Brander C, Annamalai K, et al. Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and caucasoid adults and children. *J Virol.* Jun 2000;74(12):5679-5690.

66. Sabbaj S, Bansal A, Ritter GD, et al. Cross-reactive CD8+ T cell epitopes identified in US adolescent minorities. *J Acquir Immune Defic Syndr.* Aug 1 2003;33(4):426-438.

67. Kiepiela P, Leslie AJ, Honeyborne I, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature.* Dec 9 2004;432(7018):769-775.

68. Honeyborne I, Rathod A, Buchli R, et al. Motif inference reveals optimal CTL epitopes presented by HLA class I alleles highly prevalent in southern Africa. *J Immunol.* Apr 15 2006;176(8):4699-4705.

69. Leslie A, Price DA, Mkhize P, et al. Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. *J Immunol.* Oct 1 2006;177(7):4699-4708.

70. Matthews PC, Adland E, Listgarten J, et al. HLA-A*7401-Mediated Control of HIV Viremia Is Independent of Its Linkage Disequilibrium with HLA-B*5703. *J Immunol.* Apr 15 2011.

71. Buranapraditkun S, Hempel U, Pitakpolrat P, et al. A novel immunodominant CD8+ T cell response restricted by a common HLA-C allele targets a conserved region of Gag HIV-1 clade CRF01_AE infected Thais. *PLoS ONE.* 2011;6(8):e23603.

72. Holdsworth R, Hurley CK, Marsh SG, et al. The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. *Tissue Antigens.* Feb 2009;73(2):95-170.

73. O'Brien SJ, Gao X, Carrington M. HLA and AIDS: a cautionary tale. *Trends in molecular medicine.* Sep 2001;7(9):379-381.

74. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19(6):8.

75. Hasan Z, Carlson JM, Gatanaga H, et al. Minor contribution of HLA class I-associated selective pressure to the variability of HIV-1 accessory protein Vpu. *Biochem Biophys Res Commun.* May 4 2012;421(2):291-295.

76. Parker KC, Bednarek MA, Hull LK, et al. Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *J Immunol.* Dec 1 1992;149(11):3580-3587.

77. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature.* May 23 1991;351(6324):290-296.

78. Schuler MM, Nastke MD, Stevanovikc S. SYFPEITHI: database for searching and T-cell epitope prediction. *Methods Mol Biol.* 2007;409:75-93.

79. Marsh SGE, Parham P, Barber LD. *The HLA FactsBook*. London: Academic Press; 2000.

80. Miles JJ, Douek DC, Price DA. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol Cell Biol.* Mar 2011;89(3):375-387.

81. Nielsen M, Lundegaard C, Blicher T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE.* 2007;2(8):e796.

82. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med.* 2003;54:535-551.

83. Kaslow RA, Carrington M, Apple R, et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med.* Apr 1996;2(4):405-411.

84. Pereyra F, Jia X, McLaren PJ, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* Dec 10 2010;330(6010):1551-1557.

85. Kosmrlj A, Read EL, Qi Y, et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature.* May 20 2010;465(7296):350-354.

86. Almeida JR, Price DA, Papagno L, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine.* Oct 1 2007;204(10):2473-2485.

87. Horton H, Frank I, Baydo R, et al. Preservation of T cell proliferation restricted by protective HLA alleles is critical for immune control of HIV-1 infection. *J Immunol.* Nov 15 2006;177(10):7406-7415.

88. Hasan Z, Carlson JM, Gatanaga H, et al. Minor contribution of HLA class I-associated selective pressure to the variability of HIV-1 accessory protein Vpu. *Biochem Biophys Res Commun.* Apr 7 2012.

89. Snoeck J, Fellay J, Bartha I, Douek DC, Telenti A. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. *Retrovirology.* 2011;8:87.

90. Fellay J, Ge D, Shianna KV, et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* Dec 2009;5(12):e1000791.

91. Yu XG, Lichterfeld M, Chetty S, et al. Mutually exclusive T-cell receptor induction and differential susceptibility to human immunodeficiency virus type 1 mutational escape associated with a two-amino-acid difference between HLA class I subtypes. *J Virol.* Feb 2007;81(4):1619-1631.

92. Gao X, Nelson GW, Karacki P, et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med.* May 31 2001;344(22):1668-1675.

93. Kloverpris HN, Harndahl M, Leslie AJ, et al. HIV control through a single nucleotide on the HLA-B locus. *J Virol.* Aug 15 2012.

94. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med.* Jun 8 2009;206(6):1253-1272.

95. Oxenius A, Price DA, Trkola A, et al. Loss of viral control in early HIV-1 infection is temporally associated with sequential escape from CD8+ T cell responses and decrease in HIV-1-specific CD4+ and CD8+ T cell frequencies. *J Infect Dis.* Aug 15 2004;190(4):713-721.

96. Feeney ME, Tang Y, Pfafferott K, et al. HIV-1 viral escape in infancy followed by emergence of a variant-specific CTL response. *J Immunol.* Jun 15 2005;174(12):7524-7530.

97. Kaul R, Dong T, Plummer FA, et al. CD8(+) lymphocytes respond to different HIV epitopes in seronegative and infected subjects. *J Clin Invest.* May 2001;107(10):1303-1310.

98. Streeck H, Frahm N, Walker BD. The role of IFN-gamma Elispot assay in HIV vaccine research. *Nat Protoc.* 2009;4(4):461-469.

99. Yang OO. Will we be able to 'spot' an effective HIV-1 vaccine? *Trends Immunol.* Feb 2003;24(2):67-72.

100. Feeney ME, Tang Y, Roosevelt KA, et al. Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child. *J Virol.* Aug 2004;78(16):8927-8930.

101. Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med.* Feb 1997;3(2):212-217.

102. Rolland M, Heckerman D, Deng W, et al. Broad and Gag-Biased HIV-1 Epitope Repertoires Are Associated with Lower Viral Loads. *PLoS ONE.* 2008;3(1):e1424.

103. Chouquet C, Autran B, Gomard E, et al. Correlation between breadth of memory HIV-specific cytotoxic T cells, viral load and disease progression in HIV infection. *AIDS.* Dec 6 2002;16(18):2399-2407.

104. Addo MM, Yu XG, Rathod A, et al. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J Virol.* Feb 2003;77(3):2081-2092.

105. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine.* Jan 2007;13(1):46-53.

106. Huang KH, Goedhals D, Carlson JM, et al. Progression to AIDS in South Africa Is Associated with both Reverting and Compensatory Viral Mutations. *PLoS One.* 2011;6(4):e19018.

107. Saez-Cirion A, Lacabaratz C, Lambotte O, et al. HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype. *Proc Natl Acad Sci U S A.* Apr 17 2007;104(16):6776-6781.

108. Chen H, Piechocka-Trocha A, Miura T, et al. Differential neutralization of human immunodeficiency virus (HIV) replication in autologous CD4 T cells by HIV-specific cytotoxic T lymphocytes. *J Virol.* Apr 2009;83(7):3138-3149.

109. Dahirel V, Shekhar K, Pereyra F, et al. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. *Proc Natl Acad Sci U S A.* Jul 12 2011;108(28):11530-11535.

110. Chen H, Nhdlovu ZM, Liu D, et al. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol.* 2012.

111. Fischer W, Perkins S, Theiler J, et al. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nat Med.* Jan 2007;13(1):100-106.

112. Santra S, Liao HX, Zhang R, et al. Mosaic vaccines elicit CD8+ T lymphocyte responses that confer enhanced immune coverage of diverse HIV strains in monkeys. *Nat Med.* Mar 2010;16(3):324-328.

113. Barouch DH, O'Brien KL, Simmons NL, et al. Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat Med.* Mar 2010;16(3):319-323.

114. Mostowy R, Kouyos RD, Hoof I, et al. Estimating the Fitness Cost of Escape from HLA Presentation in HIV-1 Protease and Reverse Transcriptase. *PLoS Comput Biol.* May 2012;8(5):e1002525.

115. Henn MR, Boutwell CL, Charlebois P, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* Mar 2012;8(3):e1002529.

116. Brockman MA, Chopera DR, Olvera A, et al. Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. *J Virol.* Apr 11 2012.

117. Payne RP, Kloverpris H, Sacha JB, et al. Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. *J Virol.* Oct 2010;84(20):10543-10557.

118. Mendoza D, Royce C, Ruff LE, et al. HLA B*5701-positive long-term nonprogressors/elite controllers are not distinguished from progressors by the clonal composition of HIV-specific CD8+ T cells. *J Virol.* Apr 2012;86(7):4014-4018.

119. Zuniga R, Lucchetti A, Galvan P, et al. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *J Virol.* Mar 2006;80(6):3122-3125.

120. Borghans JA, Molgaard A, de Boer RJ, Kesmir C. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS ONE.* 2007;2(9):e920.

121. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *J Virol.* Mar 2002;76(5):2298-2305.

122. Sacha JB, Chung C, Rakasz EG, et al. Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. *J Immunol.* Mar 1 2007;178(5):2746-2754.

123. Sacha JB, Chung C, Reed J, et al. Pol-specific CD8+ T cells recognize simian immunodeficiency virus-infected cells prior to Nef-mediated major histocompatibility complex class I downregulation. *J Virol.* Nov 2007;81(21):11703-11712.

124. Rauch A, James I, Pfafferott K, et al. Divergent adaptation of hepatitis C virus genotypes 1 and 3 to human leukocyte antigen-restricted immune pressure. *Hepatology.* Oct 2009;50(4):1017-1029.

125. Gaudieri S, Rauch A, Park LP, et al. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. *J Virol.* Nov 2006;80(22):11094-11104.

126. Timm J, Li B, Daniels MG, et al. Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. *Hepatology.* Aug 2007;46(2):339-349.

3. An in vitro assay to assess TCR function

With assistance from Gursev Anmole and Kinson Tse.

3.1. Introduction

Cytotoxic T lymphocytes (CTLs) monitor the intracellular environment for pathogen infections by sampling peptide epitopes presented on the surface of host T-cells in complex with human leukocyte antigen (HLA) class I molecules. When a CTL recognizes a specific peptide-HLA (pHLA) ligand through its antigen-specific TCR, a signalling cascade is initiated, leading to the activation of CTL effector functions. These functions can include perforin and granzyme-mediated lysis of the target T-cell[1], secretion of cytokines and chemokines including IFNg and MIP1ß and proliferation. However, T-cell activation can also lead to immune inhibition, through cell death mediated clonal deletion and upregulation of T-cell exhaustion markers such as PD-1[2].

Following pHLA encounter, CTL elicit heterogeneous responses[3] that can include one or many effector functions. This is determined, in part, by substantial differences in the stimulation threshold needed for various CTL responses. The amount of peptide required to induce a given T-cell function, which is often referred to as “functional avidity”, is a consequence of TCR/pHLA binding affinity as well as other factors. As it is a cell-based readout, differences in the effector and target T-cell themselves and the extracellular cytokine environment can modulate the functional threshold of CTL. Peptide-HLA density on the target T-cell surface is determined by pathogen protein expression, antigen processing efficiency, and HLA protein expression. Antigen processing related proteins and HLA expression is modulated by cytokines such as IFNg[4]. On the effector CTL, TCR and CD8 co-receptor expression are not constant[5,6]. Furthermore, co-inhibitory molecules such as PD-1 and Tim-3 can block intracellular TCR signalling events and reduce CTL reactivity[7].

TCR sequence determines T-cell antigen specificity. On CTL, TCR and the CD8 co-receptor bind pHLA in a multi-partite complex, with the subsequent signalling cascade typically referred to as "signal 1". TCR signalling is crucial in the induction of CTL activity, and biophysical parameters that define TCR/pHLA binding chemistry are likely to relate to a CTL's functional avidity. Such parameters include the dissociation constant (k_d), on-rate (k_{on}), and off-rate (k_{off}), which have been measured using techniques such as surface plasmon resonance (SPR) using purified proteins[8] and tetramer decay half-life in cell culture assays. Using SPR, the affinities of TCR binding to monomeric pHLA has been measured to be in the micromolar range[9], which are weak in comparison to typical protein-protein interactions.

The mechanics of TCR signalling continues to be a highly disputed topic. This issue is important, because the biophysical parameters of TCR binding that contribute most in determining the strength of signal 1 differ according to the signalling model that is invoked. In *the segregation/redistribution model*, signalling is mediated by the geometry of the immunological synapse. Outside of the synapse, Lck kinases and CD45 phosphatases are at equilibrium, with no net phosphorylation of ITAMs (e.g. the ground state). Within the synapse, the larger CD45 protein is size-excluded, and TCR ITAMs are increasingly activated. However, only in the presence of the correct pHLA ligand will TCR remain within the synapse sufficiently long to allow Lck-mediated phosphorylation[10]. Under such a model, persistent binding of TCR yields signalling, with k_d and k_{on} possibly representing biologically important quantities. The *serial triggering model* attempts to reconcile the observation that several hours of signalling is often necessary for T-cell activation[11], however the affinity of TCR is too low for this to occur in a single binding event. Under such a model, TCR bind and dissociate from specific pHLA in series, requiring a binding half-life long enough to allow signalling, but rapid enough to allow subsequent pHLA interactions with adjacent TCR molecules[12]. In support of this model, in a mouse immunization experiment with adoptively transferred T-cells and different peptide agonists, ligands with intermediate half lives (k_{off}) generated the strongest *in vivo* response, not the ligand with the highest affinity and longest half life[13,14].

Theoretical considerations of cellular immunity taking into account the number of T-cells in the body and the set of all possible non-self epitopes suggest that TCR binding must be highly degenerate[15]. Sewell and colleagues assayed a CTL clone against a

combinatorial peptide library, and identified epitope positions 4-6 as key TCR contact points, with their TCR having over a million possible ligands due to variability outside of this region[16].

Cross-reactivity of TCR may be important in determining the degree of CTL-mediated immune control. HIV-1 mutates rapidly due to error-prone reverse transcription, viral recombination and other factors. Different HLA-associated polymorphisms have been identified in the HIV-1 proteome, presumably representing immune escape variants. It is possible for a CTL repertoire to recognize both the wild type and the escape variant. In individuals with many TCR clonotypes recognizing the same epitope, an escape mutation might not be sufficient to escape all clonal responses. Alternatively, it is possible for an individual TCR could be sufficiently degenerate, that the escape mutant cannot interrupt TCR recognition. Thus, cross-reactivity may be an immune correlate of protection in the context of HIV infection.

Antigen sensitivity may be particularly important in mediating a rapid, robust CTL response towards HIV-1 infection[17]. HIV undergoes transcriptional latency in resting cells[18], and the HIV accessory protein Nef downregulates surface HLA expression[19]. Therefore, pHLA density may be relatively low on virus-infected cells compared uninfected tissues or another pathogen infection. Under such conditions, high antigen sensitivity may be crucial in counteracting this pathogen evasion strategy. Indeed, high avidity responses are enriched in CTL isolated from HIV elite controllers[20]. Thus, antigen sensitivity may also represent an immune correlate of protection.

The functional avidity of an immune response may have clearer vaccine design implications than a biophysical binding constant, since the phenomena being measured are cellular activities that can directly inhibit viral infection. However, the reality that antigen sensitivity is a consequence of multiple variables makes avidity challenging to study systematically, especially in the context of primary cells where many variables cannot be controlled. Here, we present a method that allows for the measurement of CTL functional avidity and cross-reactivity, while retaining control over each experimental variable that can affect the effector/target T-cell interaction. By using a cell-line approach, we are able to fix the genetic and metabolic background of both targets and effectors. Furthermore, the intrinsically molecular nature of this assay affords a degree of modularity, allowing this assay to be performed on different peptide variants,

in the presence of different molecules such as co-receptor and exhaustion markers, or by measuring different signalling pathways. Recent modelling work suggests different qualities of functional avidity such as the EC50 and E_{max} values may have a relationship to K_d, K_{on}, and K_{off}, depending on which TCR binding model is true[21]. By performing EC50 and E_{max} measurements in a genetically fixed system, we may be able to more clearly elucidate this relationship. Thus, we seek to bridge the biochemical characterizations of TCR binding with immunologically important end-points.

If particular TCR binding properties, such as peptide cross reactivity or antigen affinity, define a protective TCR clonotype, a CTL vaccine could be designed that preferentially induces the stimulation of a favourable TCR repertoire. Establishing an assay measuring these quantities is a necessary first step in such an effort.

3.2. Methods

3.2.1. Molecular cloning of TCR alpha and TCR beta genes

A CTL clone specific for the HLA-A*02:01-restricted HIV-Gag FK10 (FLGKIWPSTK) epitope was kindly provided by Dr. Mario Ostrowski (University of Toronto). Lysate was homogenized by suspending cells in RLT lysis buffer and passing the lysate through a blunt 0.8*25 mm RNAse-free syringe 10 times. Total RNA was then extracted using an RNeasy kit (Qiagen; Cat # 74104). RNA was eluted using RNAse-free water and stored at -80. 5'-RACE was performed using the SMARTer RACE cDNA amplification kit (Clontech; Cat # 634924). In brief, 125 ng of total RNA was included in the cDNA reaction, with a kit-provided adapter primer attached to the 5' end of the strand during first strand synthesis. Nested PCR was then performed using the Expand High Fidelity (Roche; Cat # 11759078001) DNA polymerase on 1:4 diluted cDNA using 3' gene specific primers that bind the constant 3'-UTR of TCR-alpha and TCR-beta, and universal primers provided by the 5'-RACE kit on the 5' end. 1st round PCR product was 1:20 diluted for 2nd round PCR. As there were two different beta UTR regions, two sets of PCR reactions were performed for the TCR beta amplification. The 1st round 3'-primers were TGTCAGGCAGTGACAAGCAG for alpha, CCTGACTGAATGGGGAGAGTCACAGGG for beta-1, and GACACTCCTGAAATGCAACCAGGCCC for beta-2. The 2nd round 3'-primers were

CAGCAGTGTTTGGCAGCTCT for alpha, AGATTTCAGCCGTGAGTGTGCAGG for beta-1, and GGAACACAGATTGGGAGCAGGTACAGGAG for beta-2. Touch-down PCR cycling conditions was performed as per manufacturer (Clontech) instructions. The correct sized product was then gel extracted, and cloned into the pCR4 plasmid using the Topo TA cloning kit (Invitrogen; Cat # K4575-02), and heat shocked at 42 °C for 30 seconds into chemically competent TOP10 cells. Transformed TOP10 cells were then plated onto prewarmed LB-kanamycin plates. Plasmid minipreps were performed using the Bio-Tek ENZA Plasmid Minikit I (OMEGA; Cat # D6942-02). Plasmids were stored in screw-cap tubes at +4 °C.

3.2.2. TCR sequence validation

TCR plasmids were sequenced using universal topo-pCR4 sequencing primers (T3_Reverse, T7_Forward, M13_Forward, M13_Reverse) provided by the Topo TA kit. Sanger sequencing was performed using the BigDye v3.1 kit (ABI) and data collected using an ABI 3130XL DNA analyzer. Chromatograms were examined using Sequencher v5 (Genecodes), and TCR sequences confirmed using the online tool IMGT/V-QUEST (www.imgt.org/IMGT_vquest).

3.2.3. Site-directed mutagenesis

Site-directed mutagenesis to introduce alanine mutations in the TCRα and TCRβ CDR3 regions was performed using the Quikchange II mutagenesis kit (Agilent; Cat # 200524-5). Mutagenesis primers were designed as per manufacturer's instructions for all 12 non-Alanine positions of the TCRα CDR3 region and all 11 non-Alanine positions of the TCRβ CDR3 region. Forward and reverse primers were ~47 bp in length spanning the region of interest and contained a single alanine codon mutation. Primer sequences are listed in Table 1. Each mutant plasmid was subsequently transformed into XL1 supercompetent *E. coli* cells (Agilent), streaked onto LB-Ampicillin agar plates and incubated at 37 °C for 16 hours overnight. Colonies were then picked, cultured in LB-Amp broth for another 16 hours, and plasmids purified and re-sequenced to confirm the mutagenesis.

3.2.4. Generation of linear PCR templates for RNA generation

Generation of *in vitro* RNA synthesis-ready linear DNA template was performed by amplifying each TCR gene from plasmids using a forward primer that partially overlap with the T7 RNA polymerase promoters flanking the insert, in the process attaching 65 thymidines to each end. The T3-facing primer was T_{65}ATTACGCCAAGCTCAGAATTAACCC and the T7-facing primer was T_{65}CGACGGCCAGTGAATTGTAATACG. Cycling conditions were +95 °C for 30 seconds, +59 °C for 30 seconds, +72 °C for 60 seconds, for 40 cycles. Purification of PCR products was performed using the Bio-tek Cycle Pure Kit (OMEGA; Cat # D6492-02).

3.2.5. In vitro RNA transcription

In vitro generation of 5'-capped, 3'-adenylated RNA was performed using the T7 MegaScript transcription kit (Ambion; Cat # AM1334) in 50 uL reactions. All steps were performed as per manufacturer instructions, except GTP was diluted 1:4 relative to the other nucleotides, and 3.6 units (5 uL) of synthetic 5' cap analog was used (Ambion; Cat # AM 8052). Cap analog was stored in single-use aliquots at -80 °C to avoid freeze/thaws. RNA synthesis was performed at +37 °C for 3 hours. RNA was then purified (RNeasy), and resuspended at 1000 ng / uL. Different TCR transcripts were then denatured by suspension in SDS/formamide-containing RNA loading buffer at +94 °C for 5 minutes and visualized on a 1% agarose gel alongside an RNA ladder. Lane-specific background-corrected relative fluorescence was determined using the Bio-Rad ImageLab software. All RNA was then diluted down to the same concentration as determined by the relative quantitation, transferred into single-use aliquots, and stored at -80 °C.

3.2.6. Plasmids

The NF-AT promoter driven luciferase reporter plasmid pNFAT-Luc was purchased from Panomics/Affymetrix (Cat # LR0050). The human CD8 alpha expression vector pORF9-hCD8a was purchased from Invivogen (Cat # porf-hcd8a). Plasmids were heat shocked into TOP10 cells, grown in LB/Amp, and purified using the Bio-Tek Maxiprep kit (OMEGA; Cat # D6922-02).

3.2.7. Cell culture

Jurkat E6.1 cells (ATCC, TIB-152) were provided by Dr. Jonathan Choy. CEM-ss cells (NIH AIDS Reagents Program, Cat # 776) that stably express HLA-A*02:01 and HLA-B*07 (referred to as CEM/A*02:01 and CEM/B*07) were generated in the lab by Tristan Markle. All cells were cultured in RPMI-1640 (Sigma) supplemented with 100 ug / mL penicillin-streptomycin and 10% fetal bovine serum. All assays were performed on cells with >90% cell viability, as determined by Trypan exclusion. All cell-washing steps were performed by centrifugation at 1500 rpm for 10 minutes at room temperature.

3.2.8. Stimulation assay

Transfection of Jurkat T-cells was performed by electroporation using a Gene Pulser MXcell Electroporation System (Bio-Rad). For each TCR tested, 10 million cells were washed and resuspended in Opti-MEM (Invitrogen) at a concentration of 50 million cells/mL. 12 ug of TCR alpha RNA, 12 ug of TCR beta RNA, 5 ug of pORF9-hCD8a plasmid, and 10 ug of pNFAT-Luc plasmid was added to each cuvette. RNA was added last to minimize degradation. Then, 10 million Jurkat T-cells (200 uL) were added to the cuvette. The electroporation conditions were as follows: square wave, 500 Volts, 2000 Farads, for 3 milliseconds. Transfected Jurkats were immediately removed from the cuvette using transfer pipettes and placed into R10 at 1.25 million / mL. 125,000 cells (100 uL) was then transferred into a 96-well flat bottom plate and allowed to rest at +37 °C for 18 hours for TCR expression. All peptides were ordered from GeneScript and were >85% pure. Peptides were dissolved in endo-toxin free Hybri-max DMSO (Sigma), then diluted to appropriate concentrations with RPMI-1640. 17 hours after the electroporation of Jurkats, 125,000 CEM cells / sample were spun down and resuspended at 1.1 million / mL in R10. 129 uL of CEM cells were then mixed with 43 uL of RPMI-diluted peptide. CEM/peptide mixture was then left to incubate at +37 °C for 1 hour. 150 uL of CEM/peptide mixture were then transferred into the wells containing 100 uL of Jurkats and allowed to co-culture for 6 hours.

3.2.9. Luciferase assay

NFAT-luciferase signal was detected using the Steady-Glo luciferase assay system (Promega; Cat #E2520), which generates a durable luminescence signal. The

luciferin solution was prepared by mixing Steady-Glo reagent with RPMI-1640 without HEPES or phenol red (Sigma; cat # R7509) at a 1:1 ratio. 250 uL of CEM/Jurkat suspension was then transferred into a 96-well V-bottom plate and washed with RPMI-1640 (without HEPES or phenol red). Cells were then suspended in 200 uL of luciferin solution, transferred into a white, opaque, flat bottom plate, and allowed to rest for 10 minutes in the dark. Luminescence was measured on an Infinite 200 PRO plate reader (Tecan), using the following settings: 3000 ms integration time, 100 ms settle time, no signal attenuation. Two measurements were performed 10 minutes apart to check for aberrant luminescence and repeated if necessary.

3.2.10. Data analysis

Data was analyzed in Prism 5 (GraphPad). The one-site binding model $Signal = \frac{E_{\max} * [Peptide]}{EC50 + [Peptide]} + NSB * [Peptide]$ was fit to the data to provide estimates for EC50 and Emax while taking non-specific binding (NSB) into account.

3.2.11. Flow cytometry

To verify the transfection efficiency, GFP RNA and CD8 plasmid were electroporated using the same methods. Surface CD8 was stained with CD8-APC-Cy7 (BD Pharmingen; Cat # 557834). Data was collected using a Guava EasyCyte 8HT flow cytometer (Millipore). Data analysis was performed using FlowJo (TreeStar).

3.2.12. HLA/peptide binding prediction

The A*02:01 binding matrix was downloaded from the BIMAS HLA Peptide binding prediction web tool (http://www-bimas.cit.nih.gov/molbio/hla_bind). netMHC-3.2 was used to predict the HLA binding properties of different peptide variants (http://www.cbs.dtu.dk/services/NetMHC).

3.2.13. Protein modelling

Prediction of TCR alpha / TCR beta interaction was performed using COTH (http://zhanglab.ccmb.med.umich.edu/COTH). The highest confidence (Z > 2.5) model

was taken. Protein visualization was done using the Chimera molecular modelling software (http://www.cgl.ucsf.edu/chimera).

3.3. Results

3.3.1. Cloning and molecular characterization of the 5B2 TCR

We implemented a 5'-RACE RT-PCR approach that allows universal cloning of TCR alpha and beta transcripts containing unknown 5' gene sequences [22]. This method avoids PCR bias that may arise from protocols that use primer mixtures designed to bind each of the ~45 alpha and ~65 beta TCR variable domains [23]. In this approach, during first strand cDNA synthesis, the MMLV-like reverse transcriptase attaches 3-5 cytosine deoxynucleotides due to its terminal transferase activity. These cytosine molecules are used as a basis for attaching an adaptor (or *capswitch oligonucleotide*) to the 3' end of the molecule during second strand synthesis, providing sequence information for 5' primer design in subsequent PCR amplifications (Figure 3.1a).

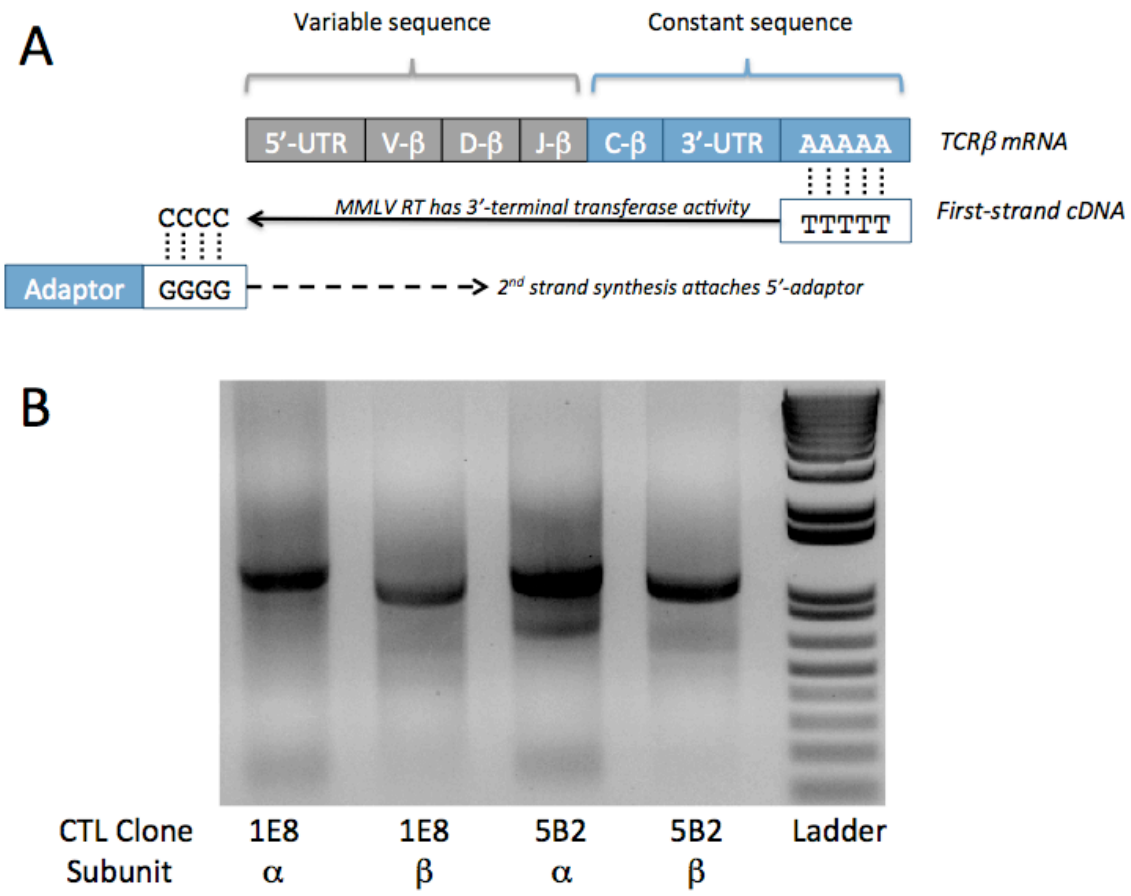

Figure 3.1 Amplification of a TCR.

A) The TCR has a constant sequence at the 3' end, but variable sequence at the 5' end. In a 5'-RACE approach, an adapter primer contains a run of G, binding the poly(C) generated by the terminal transferase activity of the MMLV RT. This adapter attaches a unique sequence to allow 5' primers to be used in subsequent PCR amplification. B) 5'-RACE amplicons of the alpha and beta subunits of two different clones: a HERV-K Env specific clone (1E8) and an HIV-Gag specific clone (5B2).

Following cDNA synthesis, nested PCR amplification was performed with primers that recognized the constant 3'-UTR region of TCR alpha and TCR beta, and universal primers which recognize the adaptor sequence upstream of the 5'-UTR. Amplicons for both TCRa and TCRb of several TCR have been successfully generated by this approach (Figure 3.1b).

Full-length TCR amplicons were cloned into pCR4 Topo/TA vector. Molecular clones were validated using gel electrophoresis, re-sequenced, and confirmed using the online tool V/QUEST (IMGT), which determines the V(D)J domains and CDR3 regions of a TCR. The FK10-specific 5B2 clone TCR was determined to have productive V(D)J

rearrangements, with a CDR3 sequence of CAVETSGTYKYIF in the alpha subunit, and CASSFGPDGYTF in the beta subunit. In summary, we can create molecular clones and genetically characterize different TCR in a streamlined, robust manner.

3.3.2. Generation of effector cells expressing exogenous TCR

Our goal was to transfect human Jurkat T-cells with the receptor proteins necessary to reprogram their antigen specificity, and then to record specific TCR/pMHC binding. We constructed an artificial Jurkat effector T-cell by transient transfection with TCR RNA, CD8a DNA, and NFAT-luciferase reporter DNA. To generate TCR RNA, we first PCR amplified each gene as a linear DNA template using a forward primer that encoded the T7 RNA polymerase promoter sequence and a reverse primer that encoded 65 thymidines (Figure 3.2). These products were used for *in vitro* T7 transcription reaction, resulting in 5' capped, 3' adenylated artificial mRNAs encoding full-length TCR alpha and TCR beta genes that were stable and translationally competent. Jurkat T-cells were transfected with the NF-AT-luciferase reporter plasmid, CD8a expression plasmid, and exogenous TCRα and TCRβ RNA by electroporation as indicated in Methods. Transfection efficiency for GFP control RNA was typically greater than 90% (Figure 3.4a), and approximately 40% for the CD8a DNA plasmid (Figure 3.4b). This discrepancy is expected, since RNA only needs to enter the cytoplasm for translation, while DNA plasmids need to enter the nucleus for transcription and subsequent protein translation. As our Jurkat T-cells express reasonable levels of exogenous TCR, coreceptor, and reporter, they are suitable for use in our functional assay described below.

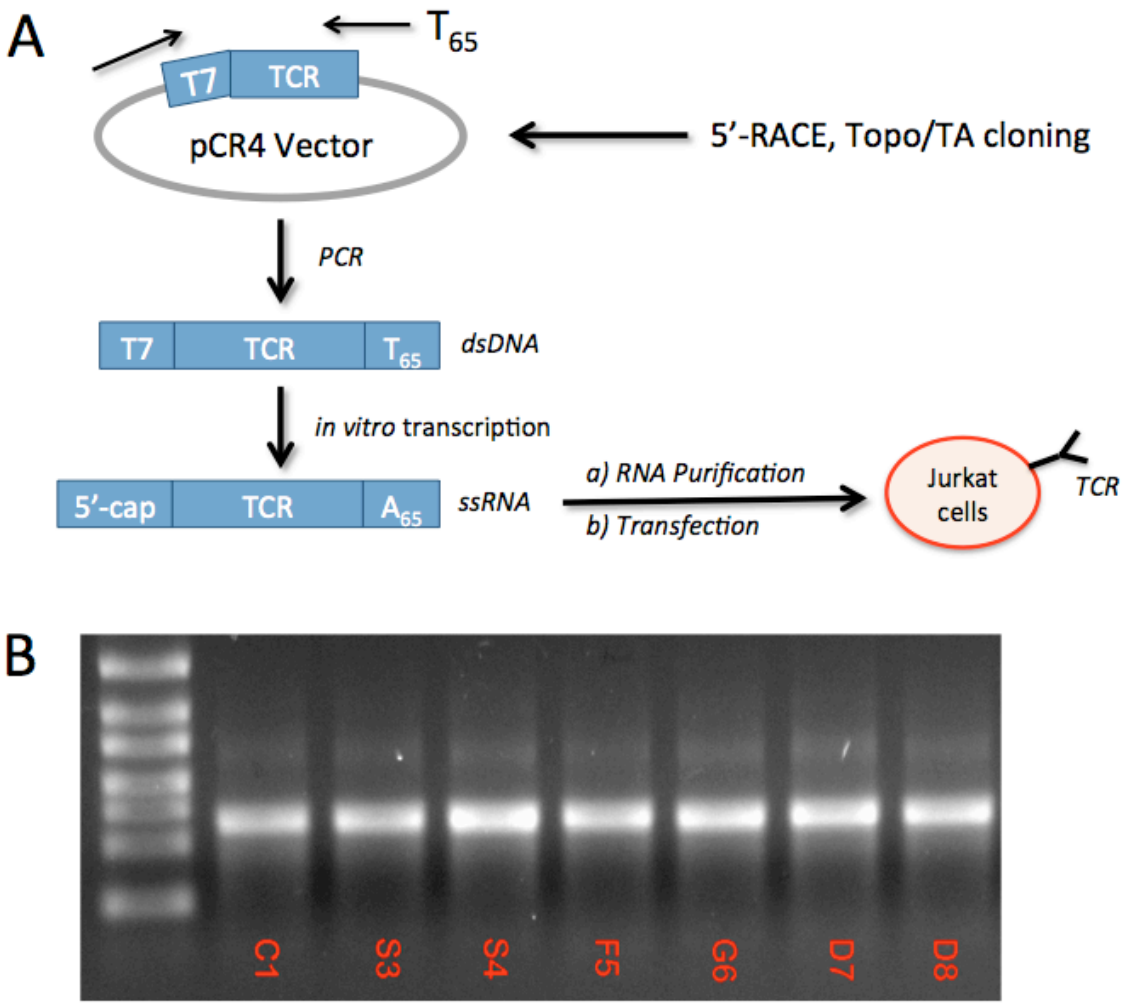

Figure 3.2 Generation of RNA from molecular plasmid clones of TCR.

A) We take advantage of PCR to attach T65 to the flanking region of the TCR insert. This allows the subsequent *in vitro* transcribed RNA to be poly-adenylated without requiring a terminal transferase reaction. B) Different *in vitro* transcription amplicons of different TCRβ CDR3 mutants. RNA was quantified by gel fluorescence and diluted to achieve the same level of relative transcript. C1 represents a cysteine to alanine mutation at the first position of the CDR3 mutant, S3 represents serine to alanine at the 3rd position, etc.

3.3.3. Sensitive, specific, and HLA-restricted recognition of A2:FK10 cells by Jurkat-TCR effector cells

To validate our assay, we examined a naturally occurring human TCR and its pHLA ligand to create a reference dataset of antigen-induced NF-AT signalling under "normal" conditions. We prepared transient TCR-expressing Jurkat effector cells by electroporation and peptide-pulsed target cells by re-suspending CEM-derived cells expressing the appropriate HLA allele for 60 minutes in medium containing various

doses of peptide. Because the test peptide displaces self-peptides bound to HLA on the surface of CEM cells, this provides a simple method to alter peptide dose. We then co-cultured Jurkat-TCR effector cells with peptide-pulsed CEM target T-cells and assessed antigen-specific and HLA-restricted recognition (Figure 3.3a).

Using the FK10-specific 5B2 TCR, we compared the peptide dose and sequence dependence of NF-AT signalling in response to FK10 versus SL9 (negative control) peptides, which are both presented by HLA A*02:01. We observed robust signalling induced by FK10 (29.3-fold relative fold induction at 30 ng / uL peptide pulse). In contrast, there was no significant signal over background for the negative control peptide (Figure 3.3b). Therefore, we concluded that transfected Jurkat effector cells expressed a functional TCR derived from CTL clone 5B2 and was specific for the FK10 peptide.

We next compared FK10 being presented by CEM-A*02+ cells versus HLA-mismatched CEM/B*07 controls, where all specific peptide induced binding was abrogated. We also checked for allogeneic responses by comparing Jurkats without target T-cells versus Jurkats cocultured with CEM effectors, and found no signal induction (data not shown). These results confirm our TCR recognition is peptide-specific, HLA-specific, and not confounded by allotypic responses.

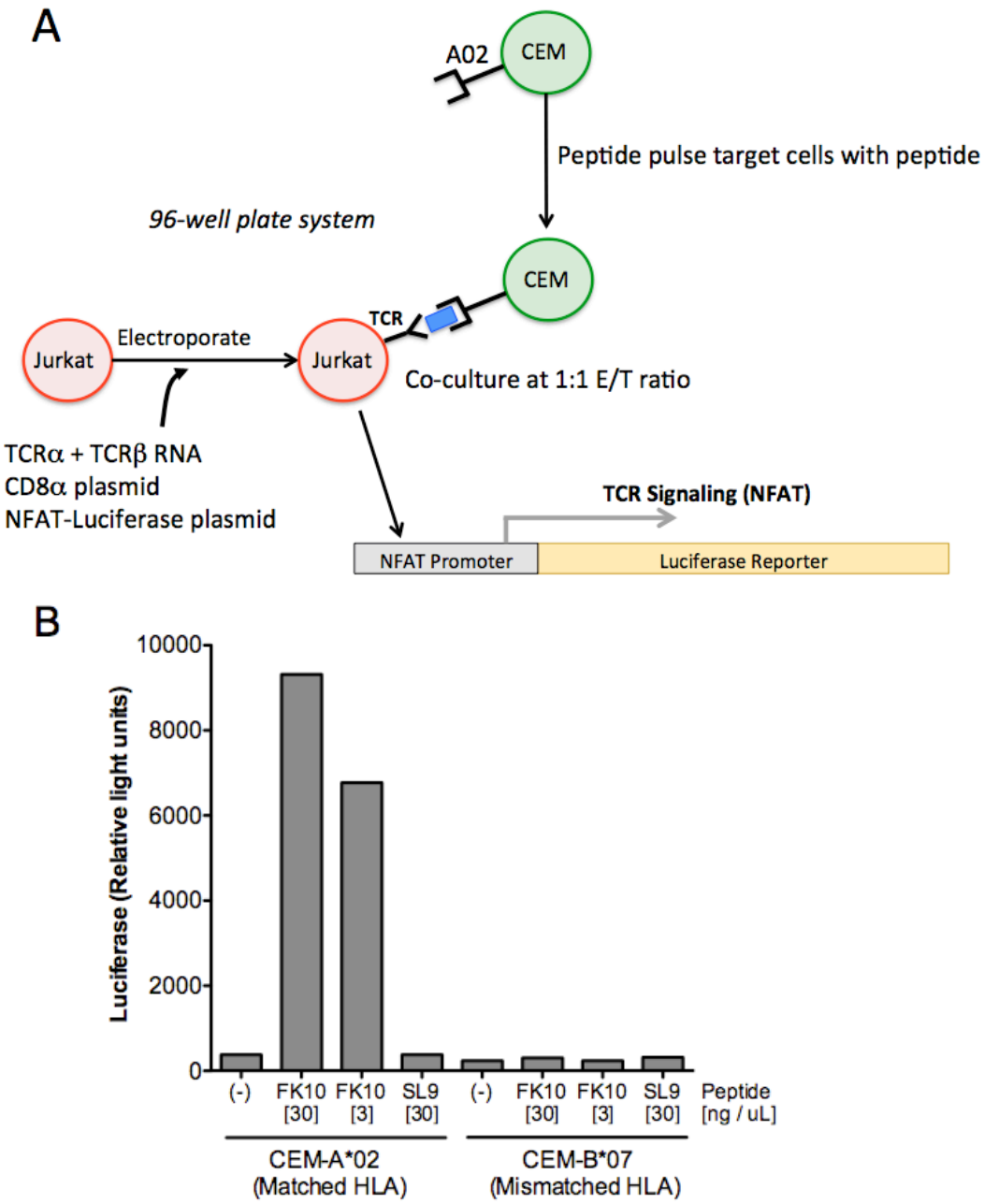

Figure 3.3 The co-culture assay.

A) TCR-transfected effectors (Jurkats), red, are stimulated by co-culture with peptide-pulsed targets (CEM), green, expressing specific HLA. The NF-AT promoter driven luciferase plasmid measures signalling from specific recognition. B) A02:01/FK10 is recognized by our TCR, but not the HLA-matched A02:01/SL9 sham peptide control. All recognition is abrogated in HLA mismatched target cell controls.

3.3.4. CD8 expression affects 5B2 TCR-mediated NF-AT signalling

The CD8 protein has been shown to enhance the sensitivity of CTL recognition by stabilizing the interaction of TCR with pHLA, and by recruiting the kinase protein Lck, which increases intracellular signalling. Indeed, CD8 can affect antigen sensitivity by up to 1 million fold [24]. Jurkat T-cells do not express CD8 endogenously, and therefore it is introduced by DNA transfection in our assay. To examine whether signalling by the FK10 TCR was sensitive to changes in CD8 expression levels, we performed coculture assays using a titration of human CD8-alpha plasmid. While some TCR signalling was observed in the absence of CD8α, the relative fold-induction of luciferase activity was greater when CD8α DNA was included in the transfection (Figure 3.4c). As the dosage of CD8a increased, the absolute amount of TCR-mediated signalling increased. However, 5 ug of CD8 DNA was observed to be optimal for the relative fold-induction between no peptide and FK10 tests, since dosages at 10 ug yielded a higher background signal. Therefore, although CD8 is not essential for 5B2 TCR recognition, the CD8 expression level on cells contributed to the relative fold-induction of signalling by altering both background as well as stimulated conditions.

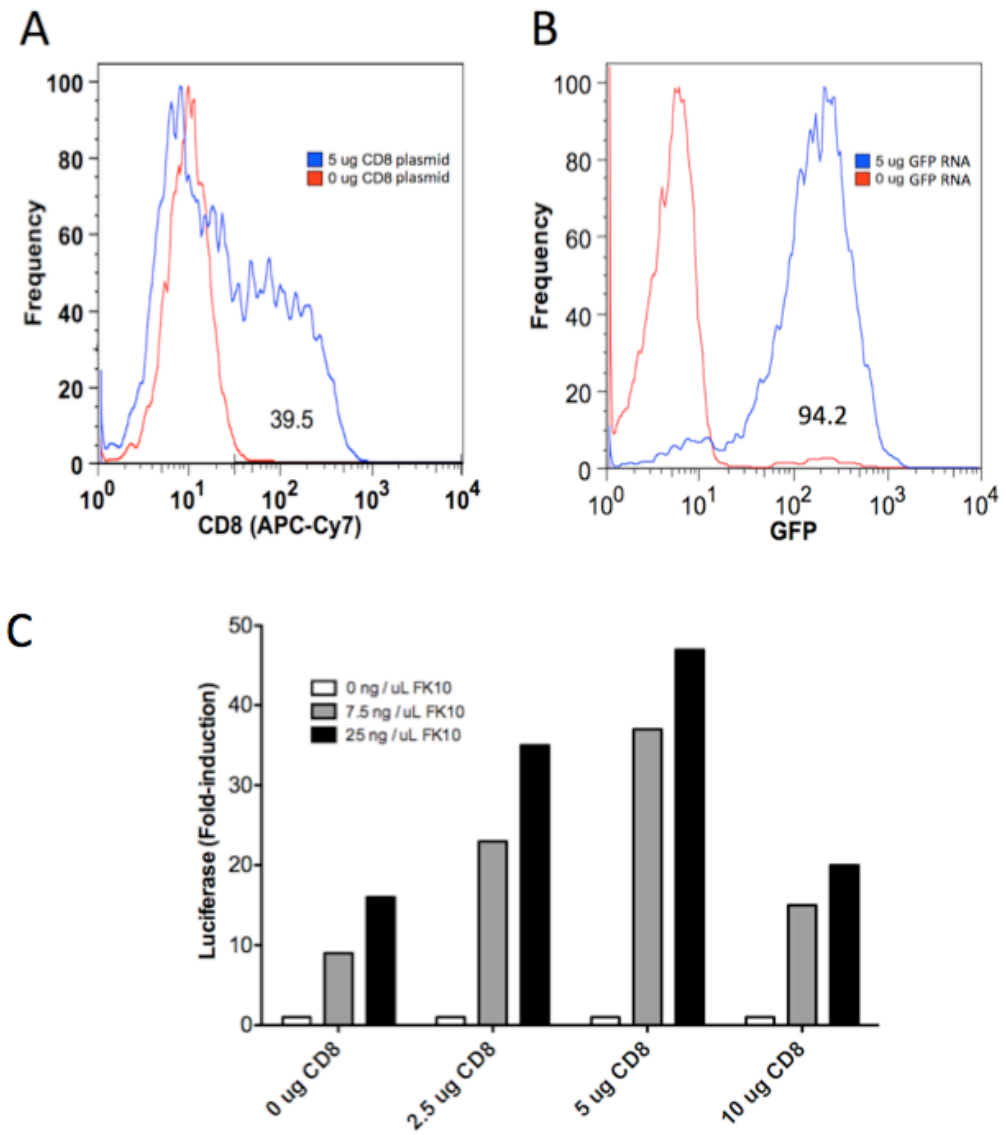

Figure 3.4 The impact of CD8 on TCR signalling.

Jurkat T-cells were transfected by electroporation with TCR RNA, GFP RNA, NFAT-luciferase DNA and various doses of CD8a DNA. At 24 hours, cells were examined by flow cytometry to assess CD8 surface expression and GFP expression, and by co-culture with FK10-pulsed CEM cells to assess TCR signalling. (A) We observed ~40% efficiency of cells expressing CD8 following DNA transfection. B) We achieved 94% efficiency of fluorescent T-cells following GFP RNA transfection. TCR-specific antibodies are unavailable. C) Jurkat effectors were transfected with different amounts of CD8 to assess the degree of CD8-dependence in this epitope. FK10-pulsed CEM/A*02:01 targets were always recognized, but optimal fold-induction of NF-AT signal was observed when Jurkat T-cells were transfected with 5 ug of CD8a DNA.

3.3.5. FK10 truncations and alanine variants reveal a minimal FS8 epitope and the importance of peptide position 4 for 5B2 TCR function

The HIV Gag FK10 peptide (FLGKIWPSYK) was originally identified as an HLA-A*02 restricted 10-mer epitope[25]; however, the predicted binding affinity of this sequence for A*02 is weak. To examine whether shorter FK10 variants could bind A*02 and be recognized by the 5B2 TCR as a minimal epitope, we performed co-culture assays with 9-mer and 8-mer peptide truncations. We observed strong signalling by Jurkat-TCR cells in both C-terminal truncated peptide (FS8 and FY9, named according to the convention: "first" AA, "last" AA, length). In contrast, no evidence of TCR signalling was observed for other truncations tested. This result demonstrates that a minimal A*02 8-mer epitope exists within FK10 that can be recognized by at least some FK10-restricted TCRs, including 5B2 (Figure 3.5a).

TCR binding is degenerate[16]. Therefore, we sought to determine which positions within the FK10 epitope provided the greatest contribution to TCR recognition and intracellular signalling. We pulsed a panel of CEM-A*02+ target T-cells with saturating concentrations of FK10 peptide variants, each encoding an alanine substitution at one position within the peptide. We observed that alanine substitutions located at positions 1, 2, 4, 6, and 7 abrogated TCR signalling (Fig 3.5b), while substitutions at other positions had little or no impact on TCR recognition. Since peptide pulsing of target T-cells circumvents intracellular antigen processing, our results suggest that these positions are crucial for either HLA binding or TCR recognition.

NetMHC is an online bioinformatic tool that predicts peptide binding to many HLA-A and HLA-B alleles with > 0.8 sensitivity and > 0.9 specificity[26]. Using this tool, the 10-mer FK10 sequence is predicted to display very weak A*02 binding (Estimated K_d of 2550 nM): however the minimal 8-mer FS8 sequence identified in our assays is predicted to display stronger A*02 binding (estimated K_d of 192 nM). Analysis of the 5 alanine variants that demonstrated poor 5B2 TCR signalling activity in our assay indicated that substitutions F1A, L2A, and W6A are likely to impair MHC binding; while P7A may impair HLA binding modestly. Notably, K4A is unlikely to affect peptide binding to HLA-A*02, and in fact was predicted to bind better than WT FS8 (Figure 3.5c). Therefore, we suspect that position 4 of the FK10 peptide represents a crucial TCR contact site, and not an HLA binding position.

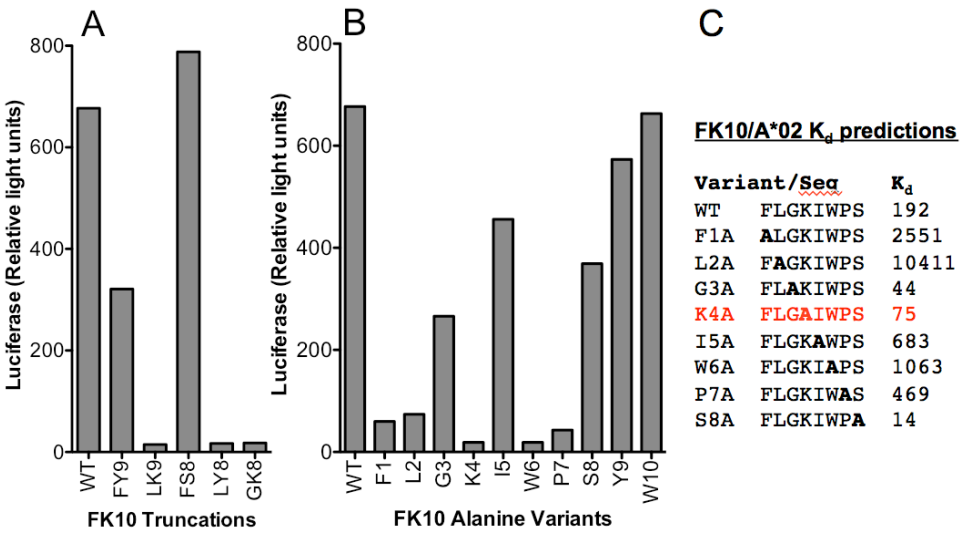

Figure 3.5 Detailed mapping of FK10 using peptide truncations and alanine variants.

A) FS8 is contained within both FY9 and FK10, and shows the strongest signal, suggesting the minimal epitope is the 8-mer FS8. B) Alanine scanning shows a total abrogation of luciferase induction at positions 1, 2, 4, 6, and 7, suggesting these are key HLA or TCR contacts. C) netMHC-3.2 estimations of peptide/HLA binding suggest mutating position 4 does not reduce HLA binding.

3.3.6. Recognition of natural FK10 variant sequences by 5B2 TCR

We next wanted to explore the potential impact of naturally occurring FK10 epitope variants on both HLA binding and TCR recognition. To do this, we performed a sequence analysis of 800 chronically HIV-infected patients to identify sequence variation within the FK10 epitope. We identified 1 variant at position 4 (K436R) and 4 variants at position 9 (H441Y, H441S, H441N, H441Q) at a frequency of 1% or greater in this cohort. None of these polymorphisms showed a significant enrichment in persons expressing A*02:01 in particular (Figure 3.6a).

We synthesized these natural peptide variants and examined their ability to activate 5B2 TCR in our Jurkat assay. None of the position 9 variants affected TCR signalling; however, the K436R mutation at position 4 abrogated luciferase activity (Figure 3.6B). K436R is not predicted to alter peptide binding to A*02 (Estimated K_d 2550 nM), suggesting that this loss of activity was instead due to a reduction in TCR recognition. Since Lysine (K) and Arginine (R) are similar positively charged amino acid residues, this result indicates that the 5B2 TCR may be highly sensitive to amino acids

at position 4 in the peptide. This conclusion is consistent with the loss of TCR signalling that was observed when alanine was substituted at position 4 (Figure 3.5b).

A

FK10 Variant	A02+	A02-
FLGKIWPSH	79%	81%
FLGKIWPSY	6.1%	5.7%
FLGKIWPSS	3.7%	2.9%
FLGKIWPSN	2.8%	3.6%
FLGKIWPSQ	1.5%	1.1%
FLGRIWPSH	0.6%	1.3%
	N=455	N=382

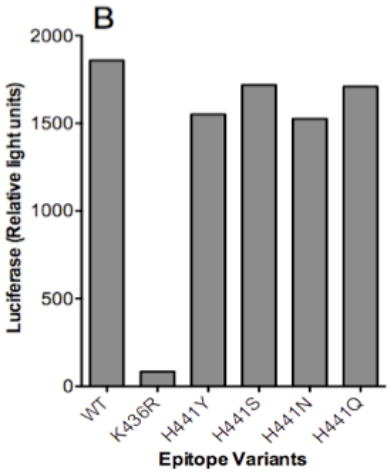

Figure 3.6 Determining the impact of HIV variants on TCR recognition

A) An analysis of chronically infected patients reveals several variants at positions 4 and 9 of the FK10 epitope, but no A*02 associated polymorphisms in particular. B) Different peptide variants were assayed for TCR recognition. No mutations at position 9 affected recognition, but the position 4 mutant K436R abrogated recognition.

3.3.7. Alanine scanning of TCR CDR3 reveal positions with differential impacts on the E_{max} and EC50

To identify functionally important positions within the 5B2 TCR sequence, we introduced alanine substitutions within the CDR3 regions of the TCRα (CAVETSGTYKYIF) and TCRβ (CASSFGPDGYTF) subunits. Mutagenesis primers are listed in table 1. For this analysis, each TCR mutant was paired with the corresponding wild type TCR chain. Jurkat T-cells expressing each receptor were stimulated by co-culture with CEM-A*02+ cells pulsed with a serial dilution of wild type FK10 peptide. While some mutations had little impact on function (Alpha: V3/S6/Y11, Beta: D8), other mutations resulted in a complete loss of TCR function (Alpha: E4/T7/Y9/K10/F13, Beta: C1/S3/F5/G6/G9/F12), and the function of many CDR3 mutants fell between these two extreme cases (Figure 3.7). Notably, among negatively charged CDR3 residues, the TCRα E4A mutant showed a complete loss of function, while the TCRß D8A showed no loss of function. We fit saturation binding curves to our data and calculated the EC50 and E_{max} values for each receptor (Table 3.1). The EC50 is the peptide concentration yielding half-maximal NFAT-luciferase signal. The E_{max} is the total luciferase signal

under saturating peptide conditions. Our data illustrate that a change in peptide affinity as measured by EC50 did not necessarily correspond with a change in E_{max}. This suggests that these values in our assay may indicate different parameters of TCR function. Additional studies will be necessary to examine this further.

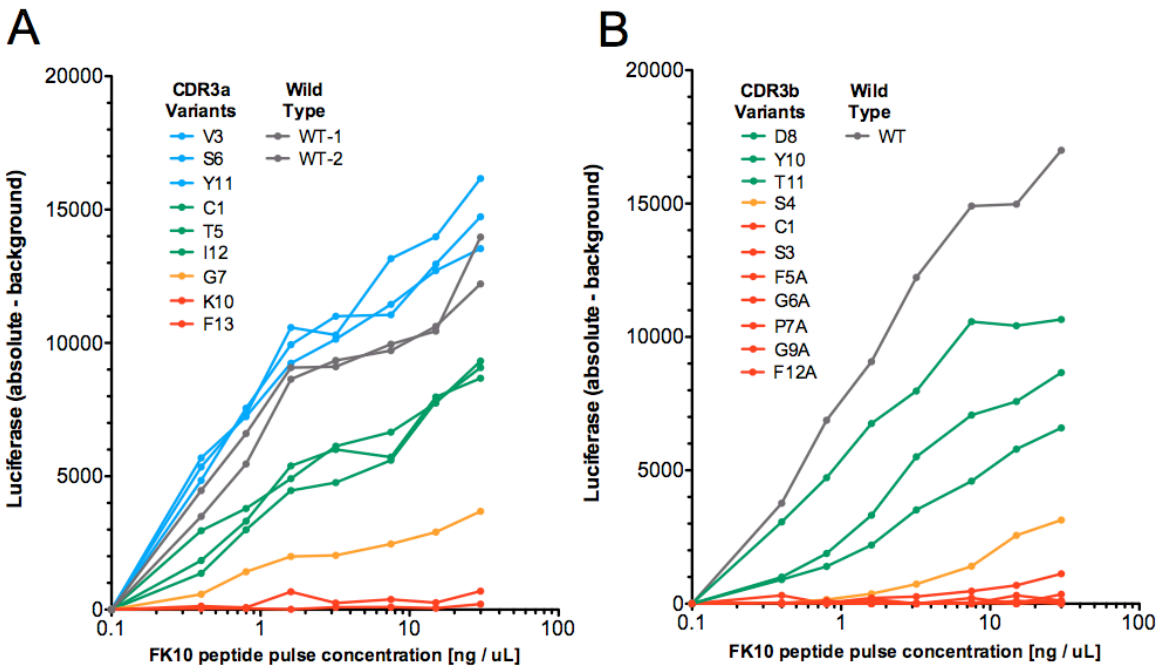

Figure 3.7 Assaying different CDR3 mutants

CDR3 mutants exhibit varying degrees of FK10 peptide induced NF-AT signalling. No impact or greater function (E_{max}) is illustrated in blue, < 50% reduction in green, > 50% reduction in orange, and total reduction in red. (A) TCRα: 3 positions (p3, p6, p11) have no impact on function, 4 positions give a partial loss (p1, p5, p7, p12), and the rest a total loss of function. B) TCRβ: 4 positions give a partial loss of function (p4, p8, p10, p11), and the rest give a total loss of function.

Table 3.1 Summary of TCRα/β CDR3 mutants

α Variant	EC50 (ng/uL)[1]	Emax (Light units)[1]	TCR activity
WT	0.77	11897	+++
V3	0.71	13718	+++
S6	0.87	15183	+++
Y11	0.63	13013	+++
C1	1.22	8325	++
T5	2.15	8836	++
G7	1.53	3393	+
I12	1.09	8567	++
E4	N.D.[2]	N.D.[2]	-
T8	N.D.[2]	N.D.[2]	-
Y9	N.D.[2]	N.D.[2]	-
K10	N.D.[2]	487	-
F13	N.D.[2]	285	-

β Variant	EC50 (ng/uL)[1]	Emax (Light units)[1]	TCR activity
WT	0.8	17468	+++
D8	0.66	11430	+++
S4	10.43	504	+
P7	13.55	2090	+
Y10	1.63	9626	++
Y11	2.17	7476	++
C1	N.D.[2]	N.D.[2]	-
S3	N.D.[2]	N.D.[2]	-
F5	N.D.[2]	N.D.[2]	-
G6	N.D.[2]	39	-
G9	N.D.[2]	271	-
F12	N.D.[2]	95	-

[1]Linear regression on log(agonist) vs response; [2]N.D. = Not determined due to poor curve fit

3.3.8. The HLA binding prediction, CDR3 variant data, and a structural model of TCR FK10 may suggest a charged interaction between p4 of FK10 and p4 of TCRα

Epitope position 4 was the only position within the FK10 peptide that abrogated TCR signalling and was not predicted to impact HLA binding. We therefore sought to explore position 4 further as a possible TCR contact site. Because this position in FK10 is a positively charged Lysine (K), we hypothesized that any corresponding residue(s) on the TCR would likely be negatively charged. Of the two negatively charged residues located in the TCRα and TCRβ CDR3 regions, only Glutamic acid (E) at position 4 in the CDR3 alpha appeared to be necessary for TCR function. Furthermore, a computational model of the 5B2 TCR structure predicts a surface pocket at the intersection of these two subunits, facing the pHLA, with E4 at its base (Figure 3.8). Therefore, our results are consistent with an interaction between a Glutamic acid residue located in the TCR alpha chain and a Lysine residue at position 4 of the FK10 peptide that is critical for TCR recognition and signalling activity. Notably, substitution of Lysine with Arginine in the FK10 peptide also failed to stimulate 5B2 TCR, suggesting that positive charge alone is not sufficient to ensure peptide recognition and TCR signalling.

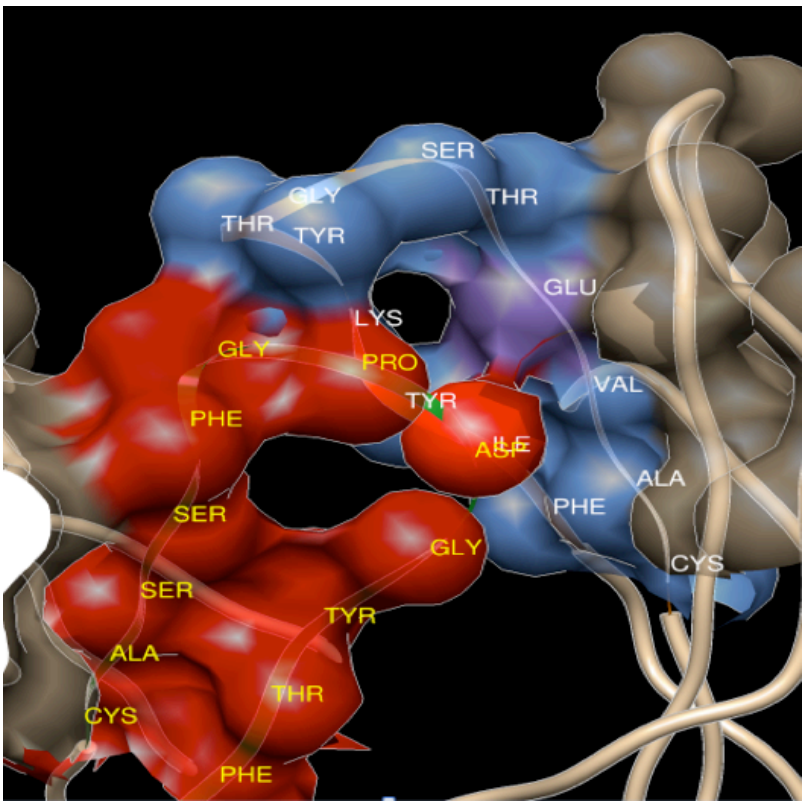

Figure 3.8 High confidence model of TCRα / TCRβ CDR3.

The only negatively charged CDR3 mutant necessary for FK10 recognition is highlighted in purple, and may represent a charged TCR / peptide interaction at the base of a binding pocket.

3.4. Discussion

The binding properties of different TCR, in the context of diverse viral variants, may play a role in determining the antiviral efficacy of a CTL response. We have implemented a technique to clone and characterize the signalling properties of a human TCR specific against the HLA-A*02 restricted HIV Gag FK10 epitope. Using 5'-RACE to amplify TCRα and TCRβ allows for the creation of molecular clones without PCR primer bias. Our TCR assay has been validated to confirm that NFAT-mediated signalling is peptide sequence specific, peptide dose dependent, and HLA-restricted. The sensitivity of our assay was sufficient to quantify the change in TCR signalling due to the effect of different peptide variants, levels of CD8 co-receptor expression, and single residue variants in the TCR alpha or beta CDR3.

NF-AT induction was enhanced by CD8 co-transfection. However, beyond a certain optimal dosage, there was an increase in non-specific signalling that reduced the relative luciferase fold-induction. However, as we were only transfecting with the CD8α subunit, which is not expected to recruit Lck[27], it is not immediately clear how surface CD8 levels contributed to non-specific TCR signalling. As CD8α interacts with HLA and colocalizes with TCR, it is possible that its overexpression may increase cell:cell interactions and/or enhance the biochemical avidity of TCR for HLA in the absence of the correct agonist peptide.

Our fine-mapping experiments indicate that FK10-specific T-cells can recognize the minimal 8-mer epitope, FS8. This is consistent with results using bioinformatics techniques that predict FK10 to be a poor binder for HLA-A*02. While it is possible that this outcome stems from inherent differences between our cell-based assay compared to the PBMC-based measurements that are more commonly used, to our knowledge the FS8 peptide has not been tested previously. In addition, the 5B2 clone TCR tested here could have unique binding properties compared to those found in the patients in which FK10 was originally fine-mapped as a 10-mer[25]. However, PBMC would be anticipated to generate a composite signal from all TCR in a patient repertoire, as opposed to clonal TCR. Additional studies to directly examine patient responses against FS8 are necessary.

Studies using FK10 alanine variants suggest that positions 1, 2, 4, 6, and 7 are essential for TCR recognition. The naturally occurring HIV variant K436R, located at position 4 in the epitope and present at low frequency in the population, also abrogated TCR signalling. By cross-referencing these results against peptide/HLA binding predictions, we postulate that FK10 position 4 is a likely TCR contact site, since neither alanine nor arginine substitutions at this site are expected to impact the HLA interaction.

Site-directed mutants in the CDR3 regions of 5B2 TCR reveal single amino acid substitutions that significantly altered TCR signalling function. In general, we observed greater changes in maximal signalling efficiency (E_{max}), as opposed to the more traditional measure of TCR functional avidity, EC50. A single negatively charged TCR mutant, E4A located in the CDR3 on TCRα, was found to completely disrupt FK10 recognition. A high-confidence structural model of this TCR positions TCRα CDR3 E4 at the base of a possible binding pocket. Thus, we hypothesize that an electrostatic

interaction between the positively charged Lysine residue K436 of FK10 and the negatively charged Glutamic acid (E4) residue in the TCRα CDR3 is critical for antigen-specific function of the 5B2 TCR. This prediction is consistent with the general observation that the TCR alpha subunit typically contacts peptide residues located in the N-terminal half of the epitope.

Moving forward, we believe this assay meets the requirements of being scalable, sensitive, and sufficiently modular to further explore the relationship between TCR/pHLA binding, T-cell signalling, and their roles in driving a protective CTL response.

Table 3.2 Site-directed mutagenesis primers

FK10 Alpha TCR CDR3 Mutagenic Primers (CDR3: CAVETSGTYKYIF)

C1A_F	CCAGTGATTC AGCCACCTAC CTCGCTGCCG TGGAGACCTC AGGAACC
C1A_R	GGTTCCTGAG GTCTCCACGG CAGCGAGGTA GGTGGCTGAA TCACTGG
V3A_F	ATTCAGCCAC CTACCTCTGT GCCGCGGAGA CCTCAGGAAC CTACAAATAC
V3A_R	GTATTTGTAG GTTCCTGAGG TCTCCGCGGC ACAGAGGTAG GTGGCTGAAT
E4A_F	CAGCCACCTA CCTCTGTGCC GTGGCGACCT CAGGAACCTA CAAATACATC
E4A_R	GATGTATTTG TAGGTTCCTG AGGTCGCCAC GGCACAGAGG TAGGTGGCTG
T5A_F	CCACCTACCT CTGTGCCGTG GAGGCCTCAG GAACCTACAA ATACATCTTTGG
T5A_R	CCAAAGATGT ATTTGTAGGT TCCTGAGGCC TCCACGGCAC AGAGGTAGGTGG
S6A_F	CCTACCTCTG TGCCGTGGAG ACCGCAGGAA CCTACAAATA CATCTTTGG
S6A_R	CCAAAGATGT ATTTGTAGGT TCCTGCGGTC TCCACGGCAC AGAGGTAGG
G7A_F	ACCTCTGTGC CGTGGAGACC TCAGCAACCT ACAAATACAT CTTTGGAACAGG
G7A_R	CCTGTTCCAA AGATGTATTT GTAGGTTGCT GAGGTCTCCA CGGCACAGAGGT
T8A_F	TCTGTGCCGT GGAGACCTCA GGAGCCTACA AATACATCTT TGGAACAGG
T8A_R	CCTGTTCCAA AGATGTATTT GTAGGCTCCT GAGGTCTCCA CGGCACAGA
Y9A_F	GTGCCGTGGA GACCTCAGGA ACCGCCAAAT ACATCTTTGG AACAGGCACC
Y9A_R	GGTGCCTGTT CCAAAGATGT ATTTGGCGGT TCCTGAGGTC TCCACGGCAC
K10A_F	CCGTGGAGAC CTCAGGAACC TACGCATACA TCTTTGGAAC AGGCACCAGG
K10A_R	CCTGGTGCCT GTTCCAAAGA TGTATGCGTA GGTTCCTGAG GTCTCCACGG
Y11A_F	TGGAGACCTC AGGAACCTAC AAAGCCATCT TTGGAACAGG CACCAGGC
Y11A_R	GCCTGGTGCC TGTTCCAAAG ATGGCTTTGT AGGTTCCTGA GGTCTCCA
I12A_F	AGACCTCAGG AACCTACAAA TACGCCTTTG GAACAGGCAC CAGGCTGAAGG
I12A_R	CCTTCAGCCT GGTGCCTGTT CCAAAGGCGT ATTTGTAGGT TCCTGAGGTCT
F13A_F	CCTCAGGAAC CTACAAATAC ATCGCTGGAA CAGGCACCAG GCTGAAGG
F13A_R	CCTTCAGCCT GGTGCCTGTT CCAGCGATGT ATTTGTAGGT TCCTGAGG

FK10 Beta TCR CDR3 Mutagenic Primers (CDR3: CASSFGPDGYTF)

C1A_F	AGGAGGACTC GGCCGTGTAT CTCGCTGCCA GCAGCTTTGG ACCAGATGGC
C1A_R	GCCATCTGGT CCAAAGCTGC TGGCAGCGAG ATACACGGCC GAGTCCTCCT
S3A_F	ACTCGGCCGT GTATCTCTGT GCCGCCAGCT TTGGACCAGA TGGCTACACC
S3A_R	GGTGTAGCCA TCTGGTCCAA AGCTGGCGGC ACAGAGATAC ACGGCCGAGT
S4A_F	CGGCCGTGTA TCTCTGTGCC AGCGCCTTTG GACCAGATGG CTACACC
S4A_R	GGTGTAGCCA TCTGGTCCAA AGGCGCTGGC ACAGAGATAC ACGGCCG
F5A_F	CCGTGTATCT CTGTGCCAGC AGCGCTGGAC CAGATGGCTA CACCTTCGG
F5A_R	CGAAGGTGTA GCCATCTGGT CCAGCGCTGC TGGCACAGAG ATACACGG
G6A_F	TGTATCTCTG TGCCAGCAGC TTTGCACCAG ATGGCTACAC CTTCGGTTCG
G6A_R	CGAACCGAAG GTGTAGCCAT CTGGTGCAAA GCTGCTGGCA CAGAGATACA
P7A_F	ATCTCTGTGC CAGCAGCTTT GGAGCAGATG GCTACACCTT CGGTTCGGGG
P7A_R	CCCCGAACCG AAGGTGTAGC CATCTGCTCC AAAGCTGCTG GCACAGAGAT
D8A_F	TCTGTGCCAG CAGCTTTGGA CCAGCTGGCT ACACCTTCGG TTCGGGGACC
D8A_R	GGTCCCCGAA CCGAAGGTGT AGCCAGCTGG TCCAAAGCTG CTGGCACAGA
G9A_F	GTGCCAGCAG CTTTGGACCA GATGCCTACA CCTTCGGTTC GGGGACCAGG
G9A_R	CCTGGTCCCC GAACCGAAGG TGTAGGCATC TGGTCCAAAG CTGCTGGCAC
Y10A_F	CCAGCAGCTT TGGACCAGAT GGCGCCACCT TCGGTTCGGG GACCAGGTTAACC
Y10A_R	GGTTAACCTG GTCCCCGAAC CGAAGGTGGC GCCATCTGGT CCAAAGCTGCTGG
T11A_F	GCAGCTTTGG ACCAGATGGC TACGCCTTCG GTTCGGGGAC CAGGTTAACC
T11A_R	GGTTAACCTG GTCCCCGAAC CGAAGGCGTA GCCATCTGGT CCAAAGCTGC
F12A_F	GCTTTGGACC AGATGGCTAC ACCGCCGGTT CGGGGACCAG GTTAACCG
F12A_R	CGGTTAACCT GGTCCCCGAA CCGGCGGTGT AGCCATCTGG TCCAAAGC

3.5. References

1. Groscurth P, Filgueira L. Killing Mechanisms of Cytotoxic T Lymphocytes. *News in physiological sciences : an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society.* Feb 1998;13:17-21.

2. Freeman GJ, Long AJ, Iwai Y, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *The Journal of experimental medicine.* Oct 2 2000;192(7):1027-1034.

3. Braciale TJ, Andrew ME, Braciale VL. Heterogeneity and specificity of cloned lines of influenza-virus specific cytotoxic T lymphocytes. *The Journal of experimental medicine.* Apr 1 1981;153(4):910-923.

4. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. *Journal of leukocyte biology.* Feb 2004;75(2):163-189.

5. Xiao Z, Mescher MF, Jameson SC. Detuning CD8 T cells: down-regulation of CD8 expression, tetramer binding, and response during CTL activation. *The Journal of experimental medicine.* Oct 29 2007;204(11):2667-2677.

6. Wooldridge L, Lissina A, Cole DK, van den Berg HA, Price DA, Sewell AK. Tricks with tetramers: how to get the most from multimeric peptide-MHC. *Immunology.* Feb 2009;126(2):147-164.

7. Sheppard KA, Fitz LJ, Lee JM, et al. PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3zeta signalosome and downstream signaling to PKCtheta. *FEBS letters.* Sep 10 2004;574(1-3):37-41.

8. Rich RL, Myszka DG. Advances in surface plasmon resonance biosensor analysis. *Current opinion in biotechnology.* Feb 2000;11(1):54-61.

9. Eisen HN, Sykulev Y, Tsomides TJ. Antigen-specific T-cell receptors and their reactions with complexes formed by peptides with major histocompatibility complex proteins. *Advances in protein chemistry.* 1996;49:1-56.

10. Davis SJ, van der Merwe PA. The kinetic-segregation model: TCR triggering and beyond. *Nature immunology.* Aug 2006;7(8):803-809.

11. Timmerman LA, Clipstone NA, Ho SN, Northrop JP, Crabtree GR. Rapid shuttling of NF-AT in discrimination of Ca2+ signals and immunosuppression. *Nature.* Oct 31 1996;383(6603):837-840.

12. Valitutti S, Lanzavecchia A. Serial triggering of TCRs: a basis for the sensitivity and specificity of antigen recognition. *Immunology today.* Jun 1997;18(6):299-304.

13. Corse E, Gottschalk RA, Krogsgaard M, Allison JP. Attenuated T Cell Responses to a High-Potency Ligand In Vivo. *PLoS biology.* Sep 2010;8(9).

14. Slansky JE, Jordan KR. The Goldilocks model for TCR-too much attraction might not be best for vaccine design. *PLoS biology.* 2010;8(9).

15. Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology today.* Sep 1998;19(9):395-404.

16. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. *The Journal of biological chemistry.* Jan 6 2012;287(2):1168-1177.

17. Almeida JR, Sauce D, Price DA, et al. Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV-suppressive activity. *Blood.* Jun 18 2009;113(25):6351-6360.

18. Coiras M, Lopez-Huertas MR, Perez-Olmeda M, Alcami J. Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nature reviews. Microbiology.* Nov 2009;7(11):798-812.

19. Blagoveshchenskaya AD, Thomas L, Feliciangeli SF, Hung CH, Thomas G. HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway. *Cell.* Dec 13 2002;111(6):853-866.

20. Mothe B, Llano A, Ibarrondo J, et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PloS one.* 2012;7(1):e29717.

21. Dushek O, Aleksic M, Wheeler RJ, et al. Antigen potency and maximal efficacy reveal a mechanism of efficient T cell activation. *Science signaling.* 2011;4(176):ra39.

22. Birkholz K, Hofmann C, Hoyer S, et al. A fast and robust method to clone and functionally validate T-cell receptors. *Journal of immunological methods.* Jul 31 2009;346(1-2):45-54.

23. Kalams SA, Johnson RP, Trocha AK, et al. Longitudinal analysis of T cell receptor (TCR) gene usage by human immunodeficiency virus 1 envelope-specific cytotoxic T lymphocyte clones reveals a limited TCR repertoire. *The Journal of experimental medicine.* Apr 1 1994;179(4):1261-1271.

24. Holler PD, Kranz DM. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. *Immunity.* Feb 2003;18(2):255-264.

25. Yu XG, Shang H, Addo MM, et al. Important contribution of p15 Gag-specific responses to the total Gag-specific CTL responses. *AIDS.* Feb 15 2002;16(3):321-328.

26. Nielsen M, Lundegaard C, Blicher T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PloS one.* 2007;2(8):e796.

27. Irie HY, Ravichandran KS, Burakoff SJ. CD8 beta chain influences CD8 alpha chain-associated Lck kinase activity. *The Journal of experimental medicine.* Apr 1 1995;181(4):1267-1273.

4. Concluding remarks

Host HLA class I allele expression is associated with different clinical outcomes following HIV infection[1]. HLA binding properties determine which peptide epitopes are presented to cytotoxic T lymphocytes (CTL), and as such, HLA contributes significantly to the generation of CTL responses in an individual. It has been proposed that HLA-dependent differences in the CTL response, such as the breadth[2] and specificity[3] of viral epitope recognition, lead to variation in the ability of these cells to suppress HIV replication and to subsequently control infection. However, mechanisms of CTL-mediated control and correlates of protection against HIV disease progression that may help to direct the design of novel vaccines and therapies remain to be clarified.

In my thesis, I have used statistically rigorous techniques[4] to analyze a large cohort of HIV+ individuals for whom paired HIV and host HLA genotype data were available. This approach allowed me to identify over 2100 HLA-associated viral polymorphisms, which highlight the evolutionary pathways used by HIV to evade host CTL. These methods were developed to avoid being confounded by viral phylogeny and HLA linkage disequilibrium. The results of my studies provide a comprehensive list of HLA-associated viral polymorphisms that can be used to examine aspects of HLA-mediated CTL selection pressure on HIV. In particular, I have used this information to investigate the contribution of an escape mutation's position within or adjacent to known epitopes, its statistical strength of association with HLA, and the total number of statistically significant CTL escape pathways per protective HLA allele.

The efficacy of antiviral CTL responses may be due to the number of epitopes targeted, the location and/or conservation of these epitopes, or particular qualities of the T cells themselves. My analysis of HLA-associated polymorphisms indicates that the total number of sites under immune selection pressure is strongly associated with a reduced risk of HIV disease progression. This result suggests that a vaccine capable of eliciting CTL responses towards multiple HIV epitopes is more likely to be successful. Notably, the only T-cell vaccine candidate against HIV that has progressed to large-

scale efficacy trials in humans (the Merck Ad5 vector expressing Gag, Pol, and Nef) failed to confer protection[5]. However immunogenicity of this vaccine appeared to be poor, with 1-3 epitope-specific responses elicited in most participants, and even fewer responses among individuals who displayed anti-vector immunity as a result of prior adenovirus type 5 infection. In light of our data, it may be prudent to revisit T-cell vaccines that can generate a larger number of CTL responses. This may clarify whether the breadth of CTL responses is a general correlate of protection, independent of host HLA.

The distribution of CTL epitopes within the viral proteome may be an important factor in determining the efficacy of the CTL response. Because HLA-associated polymorphisms indicate viral positions that are under CTL-mediated selection pressure in vivo, they provide a novel strategy to examine this issue. My analysis has revealed that HLA alleles associated with polymorphisms in the HIV-Gag protein tended to be more protective. HIV-Gag, especially the capsid protein (p24), is highly conserved. The ability to target mutationally constrained regions of the virus may be a hallmark of a protective CTL response.

The locations of CTL escape mutations within or adjacent to epitopes may also provide clues to understand correlates of effective immune responses. My analysis has revealed that viral mutations associated with protective HLA alleles tend to occur more often at epitope positions that anchor the peptide to HLA. Since anchor residue mutations are likely to impair CTL recognition more effectively than other escape pathways, understanding this correlation with protection is not obvious or intuitive. We propose that, in the context of a protective HLA, viral escape mutations at non-anchor positions are insufficient to reduce the efficacy of the CTL response. This could be possible if protective HLA alleles are able to elicit more diverse TCR repertoires, resulting in CTL populations that are inherently better able to recognize viral escape mutations that do not occur at anchor residues. Consistent with this hypothesis, a recent computational analysis of the human genome has indicated that fewer HLA-B*57 and HLA-B*27 self-peptides are expressed endogenously, which may result in less negative selection of T-cells in the thymus and a therefore a broader TCR repertoire in mature CTL[6].

Different properties of antigen-specific TCR could contribute significantly to the protective effect of HLA. For example, high antigen sensitivity of a TCR clonotype may be necessary to drive a CTL response that can recognize HIV-infected cells at early stages in the viral life cycle[7]. Alternatively, TCR binding parameters could intrinsically result in different patterns of intracellular signalling, leading to a polyfunctional CTL effector response. Finally, protective TCR clonotypes could be more cross-reactive towards viral epitope variants, constraining pathways for viral escape. To properly address these questions, it is necessary to measure TCR function directly.

I have implemented an *in vitro* assay to assess TCR function by measuring intracellular signal transduction following TCR binding and recognition of peptide/HLA ligand. Specifically, a TCR of interest is transiently expressed on an "effector" Jurkat T-cell line. TCR activation is induced by co-culture of these cells with peptide-pulsed target T-cells. I have validated this system using a TCR specific for the HLA-A*02:01-restricted FK10 epitope in HIV Gag. I have demonstrated that this assay recapitulates peptide-specific and HLA-restricted TCR signalling. Furthermore, I observed that the degree of signalling depends in part on the amount of co-receptor CD8 protein expressed on the effector T-cell surface. CD8 is often located adjacent to the TCR on the cell surface and binds a conserved domain on HLA[8], therefore, it is likely to increase functional avidity by stabilizing the trimeric TCR/pHLA complex and may also constrain the TCR/pHLA interaction. The contribution of CD8 to CTL activity can vary between clones, and likely depends on properties of the TCR/pHLA interaction, so it will be important to measure the degree of CD8 dependence for additional TCRs, epitopes, and HLA molecules, and to conduct further studies to investigate potential differences in the contribution of surface CD8α, which may stabilize TCR binding, versus CD8β, which can recruit the intracellular Lck-kinase which may directly enhance downstream signalling.

Despite several decades of research, the mechanism of TCR recognition and signalling remains inadequately defined and biochemical correlates of TCR function are still debated. The unclear relationship between functional avidity (typically measured by EC50) and maximal receptor signalling (saturation activity, or E_{max}) is highlighted in my experiments. Alanine substitutions were generated across the TCR alpha and beta CDR3 in order to identify residues necessary for antigen-specific recognition and signalling. However, we observed that changes in E_{max} were largely independent from changes in EC50. To rule out the possibility that these observed differences in signalling

are merely due to differences in RNA quality or TCR expression levels, we could measure RNA integrity using capillary electrophoresis techniques[9] (Agilent Bioanalyzer or similar) and surface expression of TCR using antibodies specific to the alpha and beta proteins. Such antibodies are not commercially available, but they may be obtained from other laboratories. It is possible that different TCR will display intrinsic variation in E_{max} and EC50 activities. If so, it will be important to understand how maximal signalling and functional avidity measures correlate with protective CTL response.

One goal in designing this assay to measure TCR function was to determine whether viral polymorphisms could disrupt CTL recognition at TCR contacts. Using alanine variants of the wild type epitope, I defined positions in FK10 that were critical for TCR recognition. By cross-referencing these results against computational predictions of peptide/HLA binding[10], I identified a putative TCR contact site at position 4 in the peptide. In addition to the alanine mutation, a naturally occurring viral mutation at the same residue, K436R, also abrogated TCR recognition. As lysine (K) and arginine (R) have similar chemical properties, this residue in Gag appears to be highly constrained and may be critical for TCR binding. Additional studies will be necessary to confirm this hypothesis. Nonetheless, this observation indicates that this in vitro assay may help to identify important TCR contact sites in viral peptides and can be used to directly assess TCR cross-reactivity towards epitope variants, which could complement more traditional approaches including the use of x-ray crystallography and in silico structural models. In the future, this assay will be used to examine the in vitro function of a larger number of antiviral TCR to assess peptide cross-reactivity and to determine whether this property is associated with protective HLA or positive clinical outcome.

As demonstrated through my studies, there are several advantages to this experimental technique:

First, using cell lines as effectors and targets allows me to measure a biologically relevant T-cellular readout such as TCR-induced signalling without being confounded by potential differences in T-cell clonotype heterogeneity.

Second, this assay is amenable to molecular manipulation, For example, I can express different TCR sequences, co-receptor isoforms, co-stimulatory/co-inhibitory molecules, and also include viral proteins to study TCR function under different

physiological settings. I can introduce mutations into TCR to assess the contribution of various residues or domains to function. Finally, using different promoter-driven reporter plasmids, I can measure different TCR signalling pathways.

Third, RNA transfection allows me to generate clonal effector T-cells rapidly and to potentially modulate the level of TCR expression more easily than other approaches. For example, lentiviral transduction requires selection and propagation of transduced cells. The additional time and labour to create panels of TCR-expressing effector cells by this method prevents larger-scale studies, such as alanine-scanning mutagenesis of TCR CDR3, which was completed in the course of my experiments.

Forth, the luciferase detection used for this method is affordable, sensitive, and generates quantitative data. Unlike ELISPOT, I am capable of measuring both EC50 and E_{max}. Since the mechanism of TCR binding and signalling remains unclear, this flexibility in experimental readout may be important to bridge biochemical characterizations of TCR/ligand binding with TCR function.

Finally, this system is highly scalable, allowing it to be used to examine a large number of peptides against a single TCR or to conduct targeted assays to directly compare the antigen-specific function of multiple TCR or mutants simultaneously. This may provide a platform for relatively high-throughput peptide library screens to assess the breadth of TCR cross-reactivity.

An improved understanding of antiviral CTL responses and TCR function will be beneficial for vaccine development. If protective CTL clonotypes are associated with certain TCR sequences or epitope binding properties, a T-cell vaccine against HIV may need to elicit a particular sub-set of the peptide-specific TCR repertoire and particular immunogens that skew the T-cell repertoire towards more protective clonotypes may be necessary. Designing such an immunogen requires an unbiased knowledge of the ligand space of a particular TCR, of which the natural viral epitope is only one ligand among an estimated 500,000 potential epitopes. We have begun using this TCR assay to screen the epitope space using combinatorial peptide pools (Figure 4.1). In preliminary studies, we have identified amino acids at particular peptide positions that appear to contribute to 5B2 TCR binding. Although such an approach assumes an independent contribution of each position to binding, a putative TCR binding motif can

be reconstructed by assaying multiple combinatorial peptide pools[11] and then subtracting the known HLA binding motif from these results. If protective TCR display a different binding motif compared to non-protective TCR, immunogens might be designed to preferentially induce only the more protective clonotype. Alternatively, if an HIV epitope is associated only with rare yet protective responses, the TCR binding motif could be used to identify a more immunogenic ligand.

Position	Poor binder					Excellent binder				FK10
1	A	D	P	Y	V	H	F			F
2	N	H	K	W	Y	R	P			L
3	R	D	G	I	T	E	M	P	V	G
4	N	E	Q	F	W	C				K
5	G	I				C	V			I
6	R	I	W			N	E	P	S	W
7	C	L	W			A	P			P
8	R	E	T			C				S
9	E	S	W			K	P			H
										K

Figure 4.1 Preliminary results from a peptide library screen.

We assayed the FK10-specific 5B2 TCR against a combinatorial peptide library containing all possible 9-mers. The recognition data represents a composite of peptide-HLA binding and TCR-pHLA binding. Thus, we may be able to use these data as a basis for creating variants of the FK10 peptide, which have improved TCR signalling properties due to a combination of these factors.

There are some technical improvements to our TCR assay that remain to be explored. Although convenient, the peptide pulse approach used for my experiments bypasses the HLA-I antigen-processing pathway. Alternatively, we could transfect target T-cells with plasmid encoding an antigenic protein of interest, or transduce cells using lentiviral transduction in order to examine processing and presentation of intracellular epitopes. We might be able to increase the sensitivity of this assay through use of different signal reporters. Although NF-AT is highly specific to TCR signalling, AP-1 may have a much higher fold-induction of signalling and be suitable in a screening approach. Also, as CD3 is rate limiting in determining surface TCR expression[12], it may be possible to co-express CD3 in order to increase TCR transfection efficiency.

In conclusion, this project supported two theses. First, correlates of CTL control can be identified through a population-level analysis of HIV evolution in the context of host HLA-mediated immune pressure. Patterns of CTL escape mutations in the virus highlight the potential importance of immune targeting against more conserved regions

of the HIV proteome. Second, the properties of antiviral TCR can be measured using a robust cell-based *in vitro* assay, providing a novel platform to examine potential mechanisms of antigen-specific T cell recognition and signalling that may elucidate new correlates of CTL control.

1. O'Brien SJ, Gao X, Carrington M. HLA and AIDS: a cautionary tale. *Trends in molecular medicine.* Sep 2001;7(9):379-381.

2. Jia M, Hong K, Chen J, et al. Preferential CTL targeting of Gag is associated with relative viral control in long-term surviving HIV-1 infected former plasma donors from China. *Cell research.* May 2012;22(5):903-914.

3. Riviere Y, McChesney MB, Porrot F, et al. Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. *AIDS research and human retroviruses.* Aug 1995;11(8):903-907.

4. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology.* Nov 2008;4(11):e1000225.

5. McElrath MJ, De Rosa SC, Moodie Z, et al. HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. *Lancet.* Nov 29 2008;372(9653):1894-1905.

6. Kosmrlj A, Read EL, Qi Y, et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature.* May 20 2010;465(7296):350-354.

7. Mothe B, Llano A, Ibarrondo J, et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PloS one.* 2012;7(1):e29717.

8. Rosenstein Y, Ratnofsky S, Burakoff SJ, Herrmann SH. Direct evidence for binding of CD8 to HLA class I antigens. *The Journal of experimental medicine.* Jan 1 1989;169(1):149-160.

9. Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology.* 2006;7:3.

10. Nielsen M, Lundegaard C, Blicher T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PloS one.* 2007;2(8):e796.

11. Rubio-Godoy V, Pinilla C, Dutoit V, et al. Toward synthetic combinatorial peptide libraries in positional scanning format (PS-SCL)-based identification of CD8+ Tumor-reactive T-Cell Ligands: a comparative analysis of PS-SCL recognition by a single tumor-reactive CD8+ cytolytic T-lymphocyte clone. *Cancer research.* Apr 1 2002;62(7):2058-2063.

12. Ahmadi M, King JW, Xue SA, et al. CD3 limits the efficacy of TCR gene therapy in vivo. *Blood.* Sep 29 2011;118(13):3528-3537.